PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
A61K 14/35, 38/16, C07H 16/12, 21/04, C12N 5/10, 15/63, C12Q 1/68, G01N 33/569

A1

(11) International Publication Number: WO 97/00067

(43) International Publication Date: 3 January 1997 (03.01.97)

(21) International Application Number: PCT/US96/10375

(22) International Filing Date: 14 June 1996 (14.06.96)

(30) Priority Data:
60/000,254 15 June 1995 (15.06.95) US

(71) Applicant *(for all designated States except US)*: UNIVERSITY OF VICTORIA [CA/CA]; 3800 Finnerty Road, Victoria, British Columbia V8W 2Y2 (CA).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: NANO, Francis, E. [US/CA]; #4-245 Ontario Street, Victoria, British Columbia V8V 1N1 (CA).

(74) Agent: POLLEY, Richard, J.; Klarquist Sparkman Campbell Leigh & Whinston, L.L.P., One World Trade Center, Suite 1600, 121 S.W. Salmon Street, Portland, OR 97204-2988 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: MYCOBACTERIUM TUBERCULOSIS DNA SEQUENCES ENCODING IMMUNOSTIMULATORY PEPTIDES

(57) Abstract

Nucleotide sequences isolated from *Mycobacterium tuberculosis* are disclosed. These sequences are shown to encode immunostimulatory peptides. The invention encompasses, among other things, vaccine preparations formulated using these peptides.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

MYCOBACTERIUM TUBERCULOSIS DNA SEQUENCES ENCODING IMMUNOSTIMULATORY PEPTIDES

CROSS REFERENCE TO RELATED CASES

This application claims the benefit of U.S. Provisional Application No. 60/000,254, filed June 15, 1995, which is incorporated herein by reference.

I. BACKGROUND

A. THE RISE OF TUBERCULOSIS

Over the past few years the editors of the Morbidity and Mortality Weekly Report have chronicled the unexpected rise in tuberculosis cases. It has been estimated that worldwide there are one billion people infected with *M. tuberculosis*, with 7.5 million active cases of tuberculosis. Even in the United States, tuberculosis continues to be a major problem especially among the homeless, Native Americans, African-Americans, immigrants, and the elderly. HIV-infected individuals represent the newest group to be affected by tuberculosis. Of the 88 million new cases of tuberculosis expected in this decade approximately 10% will be attributable to HIV infection.

The emergence of multi-drug resistant strains of *M. tuberculosis* has complicated matters further and even raises the possibility of a new tuberculosis epidemic. In the U.S. about 14% of *M. tuberculosis* isolates are resistant to at least one drug, and approximately 3% are resistant to at least two drugs. *M. tuberculosis* strains have even been isolated that are resistant to all seven drugs in the repertoire of drugs commonly used to combat tuberculosis. Resistant strains make treatment of tuberculosis extremely difficult: for example, infection with *M. tuberculosis* strains resistant to isoniazid and rifampin leads to mortality rates of approximately 90% among HIV-infected individuals. The mean time to death after diagnosis in this population is 4-16 weeks. One study reported that of nine immunocompetent health care workers and prison guards infected with drug resistant *M. tuberculosis*, five died. The expected mortality rate for infection with drug sensitive *M. tuberculosis* is 0%.

The unrelenting persistence of mycobacterial disease worldwide, the emergence of a new, highly susceptible population, and the recent appearance of drug resistant strains point to the need for new and better prophylactic and therapeutic treatments of mycobacterial diseases.

B. TUBERCULOSIS AND THE IMMUNE SYSTEM

Infection with *M. tuberculosis* can take on many manifestations. The growth in the body of *M. tuberculosis* and the pathology that it induces is largely dependent on the type and vigor of the immune response. From mouse genetic studies it is known that innate properties of the macrophage play a large role in containing disease (1). Initial control of *M. tuberculosis* may also be influenced by reactive $\gamma\delta$ T cells. However, the major immune response responsible for containment of *M. tuberculosis* is via helper T cells (Th1) and to a lesser extent cytotoxic T cells (2). Evidence suggests that there is very little role for the humoral response. The ratio of responding Th1 to Th2 cells has been proposed to be involved in the phenomenon of suppression.

Th1 cells are thought to convey protection by responding to *M. tuberculosis* T cell epitopes and secreting cytokines, particularly interferon-γ, which stimulate macrophages to kill *M. tuberculosis*. While such an immune response normally clears infections by many facultative intracellular pathogens, such as *Salmonella*, *Listeria* or *Francisella*, it is only able to contain the growth of other pathogens such as *M. tuberculosis* and *Toxoplasma*. Hence, it is likely that *M. tuberculosis* has the ability to suppress a clearing immune response, and mycobacterial components such as lipoarabinomannan are thought to be potential agents of this suppression. Dormant *M. tuberculosis* can remain in the body for long periods of time and can emerge to cause disease when the immune system wanes due to age or other effects such as infection with HIV-1.

Historically it has been thought that one needs replicating *Mycobacteria* in order to effect a protective immunization. An hypothesis explaining the molecular basis for the effectiveness of replicating mycobacteria in inducing protective immunity has been proposed by Orme and co-workers (3). These scientists suggest that antigens are pinocytosed from the mycobacterial-laden phagosome and used in antigen presentation. This hypothesis also explains the basis for secreted proteins effecting a protective immune response.

Antigens that stimulate T cells from *M. tuberculosis* infected mice or from PPD-positive humans are found in both the whole mycobacterial cells and also in the culture supernatants (3, 4, 5-7, 34). Recently Pal and Horwitz (8) were able to induce partial protection in guinea pigs by vaccinating with *M. tuberculosis* supernatant fluids. Similar results were found by Andersen using a murine model of tuberculosis (9). Other studies include reference nos. 34, 12. Although these works are far from definitive they do strengthen the notion that protective epitopes can be found among secreted proteins and that a non-living vaccine can protect against tuberculosis.

For the purposes of vaccine development one needs to find epitopes that confer protection but do not contribute to pathology. An ideal vaccine would contain a cocktail of T-cell epitopes that preferentially stimulate Th1 cells and are bound by different MHC haplotypes. Although such vaccines have never been made there is at least one example of a synthetic T-cell epitope inducing protection against an intracellular pathogen (10). It is an object of this invention to provide *M. tuberculosis* DNA sequences that encode bacterial peptides having an immunostimulatory activity. Such immunostimulatory peptides will be useful in the treatment, diagnosis and prevention of tuberculosis.

II. SUMMARY OF THE INVENTION

The present invention provides DNA sequences isolated from *Mycobacterium tuberculosis*. Peptides encoded by these DNA sequences are shown to stimulate the production of the macrophage-stimulating cytokine, gamma interferon ("INF-γ"), in mice. Critically, the production of INF-γ by CD4 cells in mice has been shown to correlate with maximum expression of protective immunity against tuberculosis (11). Furthermore, in human patients with active "minimal" or "contained" tuberculosis, it appears that the containment of the disease may be attributable, at least in part, to the production of CD4 Th-1-like lymphocytes that release INF-γ (12).

Hence, the DNA sequences provided by this invention encode peptides that are capable of stimulating T-cells to produce INF-γ. That is, these peptides act as epitopes for CD4 T-cells in the immune system. Studies have demonstrated that peptides isolated from an infectious agent and which are shown to be T-cell epitopes can protect against the disease caused by that agent when administered as a vaccine (13, 10). For example, T-cell epitopes from the parasite *Leishmania major* have been shown to be effective when administered as a vaccine (10, 13-14). Therefore, the immunostimulatory peptides (T-cell epitopes) encoded by the disclosed DNA sequences may be used, in purified form, as a vaccine against tuberculosis.

As noted, the nucleotide sequences of the present invention encode immunostimulatory peptides. In a number of instances, these nucleotide sequences are only a part of a larger open reading frame (ORF) of an *M. tuberculosis* operon. The present invention enables the cloning of the complete ORF using standard molecular biology techniques, based on the nucleotide sequences provided herein. Thus, the present invention encompasses both the nucleotide sequences disclosed herein and the complete *M. tuberculosis* ORFs to which they correspond. However, it is noted that since each of the nucleotide sequences disclosed herein encodes an immunostimulatory peptide, the use of larger peptides encoded by the complete ORFs is not necessary for the practice of the invention. Indeed, it is anticipated that, in some instances, proteins encoded by the corresponding ORFs may be less immunostimulatory than the peptides encoded by the nucleotide sequences provided herein.

One aspect of the present invention is an immunostimulatory preparation comprising at least one peptide encoded by the DNA sequences presented herein. Such a preparation may include the purified peptide or peptides and one or more pharmaceutically acceptable adjuvants, diluents and/or excipients. Another aspect of the

invention is a vaccine comprising one or more peptides encoded by nucleotide sequences provided herein. This vaccine may also include one or more pharmaceutically acceptable excipients, adjuvants and/or diluents.

Another aspect of the present invention is an antibody specific for an immunostimulatory peptide encoded by a nucleotide sequence of the present invention. Such antibodies may be used to detect the present of *M. tuberculosis* antigens in medical specimens, such as blood or sputum. Thus, these antigens may be used to diagnose tuberculosis infections.

The present invention also encompasses the diagnostic use of purified peptides encoded by the nucleotide sequences of the present invention. Thus, the peptides may be used in a diagnostic assay to detect the presence of antibodies in a medical specimen, which antibodies bind to the *M. tuberculosis* peptide and indicate that the subject from which the specimen was removed was previously exposed to *M. tuberculosis*.

The present invention also provides an improved method of performing the tuberculin skin test to diagnose exposure of an individual to *M. tuberculosis*. In this improved skin test, purified immunostimulatory peptides encoded by the nucleotide sequences of this invention are employed. Preferably, this skin test is performed with one set of the immunostimulatory peptides, while another set of the immunostimulatory peptides is used to formulate vaccine preparations. In this way, the tuberculin skin test will be useful in distinguishing between subjects infected with tuberculosis and subjects who have simply been vaccinated. In this manner, the present invention may overcome a serious limitation inherent in the present BCG vaccine/tuberculin skin test combination.

Other aspects of the present invention include the use of probes and primers derived from the nucleotide sequences disclosed herein to detect the presence of *M. tuberculosis* nucleic acids in medical specimens.

A further aspect of the present invention is the discovery that a significant proportion of the immunostimulatory peptides are homologous to proteins known to be located in bacterial cell surface membranes. This discovery suggests that membrane-bound peptides, particularly those from *M. tuberculosis*, may be a new source of antigens for use in vaccine preparations.

III. BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows the deduced amino acid sequence of the full length MTB2-92 protein.

Fig. 2 shows an SDS polyacrylamide gel (12%) representing the different stages of the purification of MTB2-92 Lane 1:- Molecular weight markers (high range, GIBCO-BRL, Grand Island, NY, U.S.A.); Lane 2:- the IPTG induced crude bacterial lysate of *E. coli* JM109 containing pMAL-MTB2-92; Lane 3:- Uninduced crude bacterial lysate of *E. coli* JM109 containing pMAL-MTB2-92; Lane 4:- Eluate from the amylose-resin column containing the MBP-MTB2-92 fusion protein; Lane 5:- Eluate shown in previous lane after cutting with protease Factor Xa; Lane 6:- Eluate from the Ni-NTA column, containing MTB2-92.

IV. DESCRIPTION OF THE INVENTION

A. DEFINITIONS

Particular terms and phrases used herein have the meanings set forth below.

"Isolated". An "isolated" nucleic acid has been substantially separated or purified away from other nucleic acid sequences in the cell of the organism in which the nucleic acid naturally occurs, i.e., other chromosomal and extrachromosomal DNA and RNA. The term "isolated" thus encompasses nucleic acids purified by standard nucleic acid purification methods. The term also embraces nucleic acids prepared by recombinant expression in a host cell as well as chemically synthesized nucleic acids.

The nucleic acids of the present invention comprise at least a minimum length able to hybridize specifically with a target nucleic acid (or a sequence complementary thereto) under stringent conditions as defined below. The length of a nucleic acid of the present invention is preferably 15 nucleotides or greater in length, although a shorter nucleic acid may be employed as a probe or primer if it is shown to specifically hybridize under stringent conditions with a target nucleic acid by methods well known in the art.

"Probes" and "primers". Nucleic acid probes and primers may readily be prepared based on the nucleic acid sequences provided by this invention. A "probe" comprises an isolated nucleic acid attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. Methods for labeling and guidance in the choice of labels appropriate for various purposes are discussed, e.g., in reference nos. 15 and 16.

"Primers" are short nucleic acids, preferably DNA oligonucleotides 15 nucleotides or more in length, which are annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, then extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods known in the art.

As noted, probes and primers are preferably 15 nucleotides or more in length, but, to enhance specificity, probes and primers of 20 or more nucleotides may be preferred.

Methods for preparing and using probes and primers are described, for example, in reference nos. 15, 16 and 17. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, © 1991, Whitehead Institute for Biomedical Research, Cambridge, MA).

"Substantial similarity". A first nucleic acid is "substantially similar" to a second nucleic acid if, when optimally aligned (with appropriate nucleotide insertions or deletions) with the other nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 75%-90% of the nucleotide bases, and preferably greater than 90% of the nucleotide bases. ("Substantial sequence complementarity" requires a similar degree of sequence complementarity.) Sequence similarity can be determined by comparing the nucleotide sequences of two nucleic acids using sequence analysis software such as the Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, Madison, WI).

"Operably linked". A first nucleic acid sequence is "operably" linked with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Recombinant". A "recombinant" nucleic acid is one that has a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques.

"Stringent Conditions" and "Specific". The nucleic acid probes and primers of the present invention hybridize under stringent conditions to a target DNA sequence, e.g., to a full length *Mycobacterium tuberculosis* gene that encodes an immunostimulatory peptide.

The term "stringent conditions" is functionally defined with regard to the hybridization of a nucleic-acid probe to a target nucleic acid (i.e., to a particular nucleic acid sequence of interest) by the hybridization procedure discussed in Sambrook et al. (1989) (reference no. 15) at 9.52-9.55. *See also*, reference no. 15 at 9.47-9.52, 9.56-9.58; reference no. 18 and reference no. 19.

Nucleic-acid hybridization is affected by such conditions as salt concentration, temperature, or organic solvents, in addition to the base composition, length of the complementary strands, and the number of nucleotide-base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art.

In preferred embodiments of the present invention, stringent conditions are those under which DNA molecules with more than 25% sequence variation (also termed "mismatch") will not hybridize. Such conditions

are also referred to as conditions of 75% stringency (since hybridization will occur only between molecules with 75% sequence identity or greater). In more preferred embodiments, stringent conditions are those under which DNA molecules with more than 15% mismatch will not hybridize (conditions of 85% stringency). In most preferred embodiments, stringent conditions are those under which DNA molecules with more that 10% mismatch will not hybridize (i.e. conditions of 90% stringency).

When referring to a probe or primer, the term "specific for (a target sequence)" indicates that the probe or primer hybridizes under stringent conditions substantially only to the target sequence in a given sample comprising the target sequence.

"Purified" - a "purified" peptide is a peptide that has been extracted from the cellular environment and separated from substantially all other cellular peptides. As used herein, the term peptide includes peptides, polypeptides and proteins. In preferred embodiments, a "purified" peptide is a preparation in which the subject peptide comprises 80% or more of the protein content of the preparation. For certain uses, such as vaccine preparations, even greater purity may be necessary.

"Immunostimulatory" - the phrase "immunostimulatory peptide" as used herein refers to a peptide that is capable of stimulating INF-γ production in the assay described in section B 5 below. In preferred embodiments, an immunostimulatory peptide is one capable of inducing greater than twice the background level of this assay determined using T-cells stimulated with no antigens or negative control antigens. Preferably, the immunostimulatory peptides are capable of inducing more than 0·01 ng/ml of INF-γ in this assay system. In more preferred embodiments, an immunostimulatory peptide is one capable of inducing greater than 10 ng/ml of INF-γ in this assay system.

B. MATERIALS AND METHODS

1. STANDARD METHODOLOGIES

The present invention utilizes standard laboratory practices for the cloning, manipulation and sequencing of nucleic acids, purification and analysis of proteins and other molecular biological and biochemical techniques, unless otherwise stipulated. Such techniques are explained in detail in standard laboratory manuals such as Sambrook et al. (15); and Ausubel et al. (16).

Methods for chemical synthesis of nucleic acids are discussed, for example, in reference nos. 20 and 21. Chemical synthesis of nucleic acids can be performed, for example, on commercial automated oligonucleotide synthesizers.

2. ISOLATION OF *MYCOBACTERIUMTUBERCULOSIS* DNA SEQUENCES ENCODING IMMUNOSTIMULATORY PROTEINS

Mycobacterium tuberculosis DNA was obtained by the method of Jacobs et al. (22). Samples of the isolated DNA were partially digested with one of the following restriction enzymes *Hin*PI, *Hpa*II, *Aci*I, *Taq*I, *Bsa*HI, *Nar*I. Digested fragments of 0·2-5kb were purified from agarose gels and then ligated into the *Bst*BI site in front of the truncated *phoA* gene in one or more of the three phagemid vectors pJDT1, pJDT2, and JDT3.

A schematic representation of the phagemid vector pJDT2 is provided in Mdluli et al. (1995) (reference no. 31). The pJDT vectors were specifically designed for cloning and selecting genes encoding cell wall-associated, cytoplasmic membrane associated, periplasmic or secreted proteins (and especially for cloning such genes from GC rich genomes, such as the *Mycobacterium tuberculosis* genome). The vectors have a *Bst*BI cloning site in frame with the bacterial alkaline phosphatase gene (*phoA*) such that cloning of an in-frame sequence into the cloning site will result in the production of a fusion protein. The *phoA* gene encodes a version of the alkaline phosphatase that lacks a signal sequence; hence, only if the DNA cloned into the *Bst*BI site includes a signal sequence or a transmembrane sequence can the fusion protein be secreted to the medium or inserted into cytoplasmic membrane, periplasm or cell wall. Those clones encoding such fusion proteins may be detected by

plating clones on agar plates containing the indicator 5-bromo-4-chloro-3-indolyl-phosphate. Alkaline phosphatase converts this indicator to a blue colored product. Hence, those clones containing secreted alkaline phosphatase fusion proteins will produce the blue color.

The three vectors in this series (pJDT1, 2 and 3) have the *Bst*BI restriction sites located in different reading frames with respect to the *pho*A gene. This increases the likelihood of cloning any particular gene in the correct orientation and reading frame for expression by a factor of 3. Reference no. 31 describes pJDT vectors in detail.

3. SELECTION OF SECRETED FUSION PROTEINS

The recombinant clones described above were transformed into *E. coli* and plated on agar plates containing the indicator 5-bromo-4-chloro-3-indolyl-phosphate. Production of blue pigmentation, produced as a result of the action of alkaline phosphatase on the indicator, indicated the presence of secreted cytoplasmic membrane periplasmic, cell wall associated or outer membrane fusion proteins (because the bacterial alkaline phosphatase gene in the vector lacks a signal sequence and could not otherwise escape the bacterial cell). A similar technique has been used to identify *M. tuberculosis* genes encoding exported proteins by Lim et al. (32).

Those clones producing blue pigmentation were picked and grown in liquid culture to facilitate the purification of the alkaline phosphatase fusion proteins. These recombinant clones were designated according to the restriction enzyme used to digest the *Mycobacterium tuberculosis* DNA (thus, clones designated A#2-1, A#2-2 etc were produced using *Mycobacterium tuberculosis* DNA digested with *Aci*I).

4. PURIFICATION OF SECRETED FUSION PROTEINS

PhoA fusion proteins were extracted from the selected *E. coli* clones by cell lysis and purified by SDS polyacrylamide gel electrophoresis. Essentially, individual *E. coli* clones are grown overnight at 30°C with shaking in 2 ml LB broth containing ampicillin, kanamycin and IPTG. The cells are precipitated by centrifugation and resuspended in 100 μL Tris -EDTA buffer. 100 μL lysis buffer (1% SDS, 1mMEDTA, 25mM DTT, 10% glycerol and 50 mM tris-HCl, pH 7.5) is added to this mixture and DNA released from the cells is sheared by passing the mixture through a small gauge syringe needle. The sample is then heated for 5 minutes at 100°C and loaded onto an SDS PAGE gel (12 cm x 14 cm x 1.5 mm, made with 4% (w/v) acrylamide in the stacking section and 10% (w/v) acrylamide in the separating section). Several samples from each clone are loaded onto each gel.

The samples are electrophoresed by application of 200 volts to the gel for 4 hours. Subsequently, the proteins are transferred to a nitrocellulose membrane by Western blotting. A strip of nitrocellulose is cut off to be processed with antibody, and the remainder of the nitrocellulose is set aside for eventual elution of the protein. The strip is incubated with blocking buffer and then with anti-alkaline phosphatase primary antibody, followed by incubation with anti-mouse antibody conjugated with horse radish peroxidase. Finally, the strip is developed with the NEN DuPont Renaissance kit to generate a luminescent signal. The migratory position of the PhoA fusion protein, as indicated by the luminescent label, is measured with a ruler, and the corresponding region of the undeveloped nitrocellulose blot is excised.

This region of nitrocellulose, which contains the PhoA fusion protein, is then incubated in 1 ml 20% acetronitrile at 37°C for 3 hours. Subsequently, the mixture is centrifuged to remove the nitrocellulose and the liquid is transferred to a new test tube and lyophilized. The resulting protein pellet is dissolved in 100 μL of endotoxin-free, sterile water and precipitated with acetone at -20°C. After centrifugation the bulk of the acetone is removed and the residual acetone is allowed to evaporate. The protein pellet is re-dissolved in 100 μL of sterile phosphate buffered saline. This procedure can be scaled up by modification to include IPTG induction 2 hours prior to cell harvesting, washing nitrocellulose membranes with PBS prior to acetonitrile extraction and lyophilization of acetonitrile extracted and acetone precipitated protein samples.

5. DETERMINATION OF IMMUNOSTIMULATORY CAPACITY IN MICE

The purified alkaline phosphatase - *Mycobacterium tuberculosis* fusion peptides encoded by the recombinant clones were then tested for their ability to stimulate INF-γ production in mice. The test used to determine INF-γ stimulation is as essentially that described by Orme et al. (11).

Essentially, the assay method is as follows: The virulent strain *M. tuberculosis* Erdman is grown in Proskauer Beck medium to mid-log phase, then aliquoted and frozen at -70°C for use as an inoculant. Cultures of this bacterium are grown and harvested and mice are inoculated with 1 x 10^5 viable bacteria suspended in 200 μl sterile saline via a lateral tail vein on day one of the test.

Bone marrow-derived macrophages are used in the test to present the bacterial alkaline phosphatase-*Mycobacterium tuberculosis* fusion protein antigens. These macrophages are obtained by harvesting cells from mouse femurs and culturing the cells in Dulbecco's modified Eagle medium as described by Orme et al. (11). Eight to ten days later, up to ten μg of the fusion peptide to be tested is added to the macrophages and the cells are incubated for 24 hours.

The CD4 cells are obtained by harvesting spleen cells from the infected mice and then pooling and enriching for CD4 cells by removal of adherent cells by incubation on plastic Petri dishes, followed by incubation for 60 minutes at 37°C with a mixture of J11d.2, Lyt-2.43, and GL4 monoclonal antibody (mAb) in the presence of rabbit complement to deplete B cells and immature T cells, CD8 cells, and γδ cells, respectively. The macrophages are overlaid with 10^6 of these CD4 cells and the medium is supplemented with 5 U IL-2 to promote continued T cell proliferation and cytokine secretion. After 72 hours, cell supernatants are harvested from sets of triplicate wells and assayed for cytokine content.

Cytokine levels in harvested supernatants are assayed by sandwich ELISA as described by Orme et al. (11).

6. DETERMINATION OF IMMUNOSTIMULATORY CAPACITY IN HUMANS

The purified alkaline phosphatase - *Mycobacterium tuberculosis* fusion peptides encoded by the recombinant clones or by synthetic peptides are tested for their ability to induce INF-γ production by human T cells in the following manner.

Blood from tuberculin positive people (producing a tuberculin positive skin test) is collected in EDTA coated tubes, to prevent clotting. Mononuclear cells are isolated using a modified version of the separation procedure provided with the NycoPrep™ 1.077 solution (Nycomed Pharma AS, Oslo, Norway). Briefly, the blood is diluted in an equal volume of a physiologic solution, such as Hanks Balanced Salt solution (HBSS), and then gently layered over top of the Nycoprep solution in a 2 to 1 ratio in 50 ml tubes. The tubes are centrifuged at 800 x g for 20 minutes and the mononuclear cells are then removed from the interface between the Nycoprep solution and the sample layer. The plasma is removed from the top of the tube and filtered through a 0.2 micron filter and is then added to the tissue culture media. The mononuclear cells are washed twice: the cells are diluted in a physiologic solution, such as HBSS or RPMI 1640, and centrifuged at 400 x g for 10 minutes. The mononuclear cells are then resuspended to the desired concentration in tissue culture media (RPMI 1640 containing 10% autologous serum, Hepes, non-essential amino acids, antibiotics and polymixin B). The mononuclear cells are then cultured in 96 well microtitre plates.

Peptides or PhoA fusion proteins are then added to individual wells in the 96 well plate, and cells are then placed in an incubator (37°C, 5% CO_2). Samples of the supernatants (tissue culture media from the wells containing the cells) are collected at various time points (from 3 to 8 days) after the addition of the peptides or PhoA fusion proteins. The immune responsiveness of T cells to the peptides and PhoA fusion proteins is assessed by measuring the production of cytokines (including gamma-interferon).

Cytokines are measured using an Enzyme Linked Immunosorbent Assay (ELISA), the details of which are described in the Cytokine ELISA Protocol in the PharMingen catalogue (PharMingen, San Diego, California). For measuring for the presence of human gamma-interferon, wells of a 96 well microtitre plate are coated with a capture antibody (anti-human gamma-interferon antibody). The sample supernatants are then added to individual wells. Any gamma-interferon present in the sample will bind to the capture antibody. The wells are then washed. A detection antibody (anti-human gamma-interferon antibody), conjugated to biotin, is added to each well, and will bind to any gamma-interferon that is bound to the capture antibody. Any unbound detection antibody is washed away. An avidin peroxidase enzyme is added to each well (avidin binds tightly to the biotin on the detection antibody). Any excess unbound enzyme is washed away. Finally, a chromogenic substrate for the enzyme is added and the intensity of the colour reaction that occurs is quantitated using an ELISA plate reader. The quantity of the gamma-interferon in the sample supernatants is determined by comparison with a standard curve using known quantities of human gamma-interferon.

Measurement of other cytokines, such as Interleukin-2 and Interleukin-4, can be determined using the same protocol, with the appropriate substitution of reagents (monoclonal antibodies and standards).

7. DNA SEQUENCING

The sequencing of the alkaline phosphatase fusion clones was undertaken using the AmpliCycle thermal sequencing kit (Perkin Elmer, Applied Biosystems Division, 850 Lincoln Centre Drive, Foster City, CA 94404, U.S.A.), using a primer designed to read out of the alkaline phosphatase gene into the *Mycobacterium tuberculosis* DNA insert, or primers specific to the cloned sequences.

C. RESULTS

1. IMMUNOSTIMULATORY CAPACITY

More than 300 fusion clones were tested for their ability to stimulate INF-γ production. Of these, 80 clones were initially designated to have some ability to stimulate INF-γ production. Tables 1 and 2 show the data obtained for these 80 clones. Clones placed in Table 1 showed the greatest ability to stimulate INF-γ production (greater than 10 ng/ml of INF-γ) while clones placed in Table 2 stimulated the production of between 2 ng/ml and 10 ng/ml of INF-γ. Background levels of INF-γ production (i.e., levels produced without any added *M. tuberculosis* antigen) were subtracted from the levels produced by the fusions to obtain the figures shown in these tables.

TABLE 1

Immunostimulatory AP-fusion clones

No.	Name	INF	Fus-MW	TBport	coding	Similarity (score)
1	AciI#1-152	>40,000	~65,000	~23,400	~633	*M. avium* acetolactate synthase (98^+)
2	AciI#1-247	>40,000	~160,000	~118,400	~3,198	peptide synthetase (153)
3	AciI#1-264	>40,000	~72,500	~30,900	~833	nothing evident
4	AciI#1-435	>40,000	~80,000	~38,400	~1,038	*M. smegmatis* ethambutol resistance gene EmbA (624)

TABLE 1

Immunostimulatory AP-fusion clones

No.	Name	INF	Fus-MW	TBport	coding	Similarity (score)
5	HinP#1-27	>20,000	59,000	17,400	471	nothing evident
6	HinP#2-92	>20,000	74,600	33,000	891	1. *M. tuberculosis* ORF MTCY190.11C (1794+) 2. Cytochrome C oxidase subunit II (141)
7	HinP#2-145	>20,000	60,000	13,900	375	nothing evident
8	HinP#2-150	>20,000	55,000	13,400	362	nothing evident
9	HinP#1-200	>20,000	53,500	11,900	321	nothing evident
10	HinP#3-30	>20,000	69,000	27,400	740	*M. leprae* chromosome sequence in B983 region (281+)
11	AciI#2-2	>20,000	70,000	28,400	768	*M. leprae* chromosome sequence within region B1529 (139)
12	AciI#2-23	>20,000	75,000	33,400	903	Region within sequence MD0009 of the *M. leprae* chromosome
13	AciI#2-506	>20,000	60,000	18,400	498	nothing evident
14	AciI#2-511	>20,000	~60,000	~18,400	~498	nothing evident
15	AciI#2-639	>20,000	~60,000	~18,400	~498	nothing evident
16	AciI#2-822	>20,000	~45,000	~3,400	~93	*M. tuberculosis* sequence within region MD0074 (U27357) (551+)
17	AciI#2-823	>20,000	~46,500	~4,900	~132	nothing evident
18	AciI#2-825	>20,000	~150,000	~110,000	~2,970	*M. tuberculosis* sequence MTCY31.03c (431)
19	AciI#2-827	>20,000	~48,000	~6,400	~174	cytochrome d oxidase
20	AciI#2-898	>20,000	~49,000	~7,400	~201	nothing evident

TABLE 1

Immunostimulatory AP-fusion clones

No.	Name	INF	Fus-MW	TBport	coding	Similarity (score)
21	AciI#2-1084	>20,000	~73,000	~31,400	~849	Sequences within *M. tuberculosis* clone X68281 (96$^+$) and *M. leprae* clone B983 (122$^+$)
22	AciI#3-47	>20,000	~55,000	~13,400	~363	nothing evident
23	AciI#3-133	>20,000	~55,000	~13,400	~363	nothing evident
24	AciI#3-166	>20,000	~48,000	~6,400	~174	nothing evident
25	AciI#3-167	>20,000	~65,000	~23,400	~633	*M. leprae* DNA sequence within region B983 (588$^+$)
26	AciI#3-206	>20,000	~65,000	~23,400	~633	*M. leprae* DNA sequence within chromosome region MD0092 (91)
27	HinP#1-31	14,638	~46,000	~4,400	~120	*M. tuberculosis* 19 kDa lipo-protein antigen precursor (218)
28	HinP#1-144	13,546	~70,000	~23,900	~645	*M. leprae* DNA sequence within chromosome region B983 (78)
29	HinP#1-3	11,550	~49,000	~7,400	~200	*M. leprae* DNA sequence within chromosome region B983 (100$^+$)
30	AciI#1-486	11,416	~45,000	~3,400	~93	nothing known
31	AciI#1-426	11,135	~47,500	~5,900	~160	Dipeptide transport protein (65)
32	AciI#2-916	10,865	~75,000	~33,400	~903	nothing evident

Abbreviations: INF: pg/ml of INF-γ produced using fusion to stimulate immune T-cells. Fus. MW: Relative molecular weight of the fusion protein in Da. TB port.: Estimated amount of fusion attributable to the *M. tuberculosis* protein. Coding: Amount of DNA needed to encode TB portion of fusion proteins (in base pairs). Similarity: Amino acid sequence similarity seen by analysis of DNA via the BLASTX or TBLASTX* programs. Scores for alignments are indicated in (). Due to the high G+C nature of M. TB DNA many false positives are evident. Only scores above 100 have good credibility.

TABLE 2

Immunostimulatory AP-fusion clones (cont'd)

No.	Clone Name	INF	Fus-MW	TBport	coding	Similarity (score)
1	AciI#1-62	3,126	~43,000	~1,400	~39	*M. tuberculosis* MTCY 190.11C cytochrome C oxidase subunit II (198) *M. leprae* sequence in B1551 region (1087+)
2	AciI#2-14	6,907	~45,000	~3,400	~93	nothing evident
3	AciI#2-26	3,089	~72,000	~30,400	~822	nothing evident
4	AciI#2-35	3,907	~45,000	~3,400	~93	Possibly similar to *M. leprae* sequence in the B983 region (116+)
5	AciI#2-147	5,464				nothing evident
6	AciI#2-508	7,052	~70,000	~28,400	~768	Similar to sequence of the *M. leprae* ORF encoding gp U00018 (125) and similar to sequence in the B2168 c2-209 region of *M. leprae* genome (225+)
7	AciI#2-510	2,445	~69,000	~27,400	~741	nothing evident
8	AciI#2-523	2,479	~50,000	~8,400	~228	Similar to *M. tuberculosis* sequence z70692 from clone Y427 (96)
9	AciI#2-676	3,651	~70,000	~28,400	~768	Similar to AciI#2-639
10	AciI#2-834	5,942	~60,000	~13,900	~375	nothing evident
11	AciI#2-854	5,560	~44,000	~2,400	~66	nothing evident
12	AciI#2-872	2,361	~47,000	~5,400	~147	nothing evident
13	AciI#2-874	2,171	~45,000	~3,400	~93	nothing evident
14	AciI#2-8841	2,729	~85,000	~43,400	~1173	Isocitrate dehydrogenase (247)
15	AciI#2-894	3,396	~70,000	~28,400	~768	nothing evident
16	AciI#2-1014	6,302	~45,000	~3,400	~93	nothing evident
17	AciI#2-1018	4,642	~55,000	~13,400	~363	nothing evident
18	AciI#2-1025	3,582	~45,000	~3,400	~93	nothing evident
19	AciI#2-1034	2,736	~80,000	~38,400	~103	nothing evident

TABLE 2

Immunostimulatory AP-fusion clones (cont'd)

No.	Clone Name	INF	Fus-MW	TBport	coding	Similarity (score)
20	AciI#2-1035	3,454	~46,000	~4,400	~120	nothing evident
21	AciI#2-1089	8,974	~65,000	~23,400	~633	Similar to *M. tuberculosis* sequence X75361 and sequence in *M. bovis* MD0057 and U34849 regions. Immunogenic proteins MPB64 and MPT64 are homologous.
22	AciI#2-1090	7,449	~65,000	~23,400	~633	nothing evident
23	AciI#2-1104	5,148	~68,000	~26,400	~714	Similar to *M. tuberculosis* sequence X80268 and to cds 1 (256) in *M. leprae* sequence region MD0045 (169$^+$); secreted antigenic protein.
24	AciI#3-9	3,160	~67,000	~25,400	~687	nothing evident
25	AciI#3-12	3,891	~75,000	~33,400	~903	Penicillin binding protein; similar to *M. leprae* sequence within genomic clone B1529
26	AciI#3-15	4,019	~65,000	~23,400	~633	nothing evident
27	AciI#3-21	2,301	~69,000	~27,400	~741	nothing evident
28	AciI3-78	2,905	~65,000	~23,400	~633	Similar to sequence within *M. leprae* genomic clone B983
29	AciI#3-134	3,895	~45,000	~3,400	~93	nothing evident
30	AciI#3-204	4,774	~60,000	~13,900	~375	nothing evident
31	AciI#3-214	7,333	~50,000	8,400	~228	nothing evident
32	AciI#3-243	2,857	~65,000	~23,400	~633	nothing evident
33	AciI#3-281	2,943	~65,000	~23,400	~633	Similar to sequence within *M. leprae* genomic clone B983
34	Bsa HI#1-21	8,122	~90,000	~48,400	~1,209	nothing evident
35	HinP#1-12	2,905	~66,000	~24,400	~660	possible tyrosine phosphatase

TABLE 2

Immunostimulatory AP-fusion clones (cont'd)

No.	Clone Name	INF	Fus-MW	TBport	coding	Similarity (score)
36	HinP#2-23	2,339	~43,000	~1,400	~39	Similar to sequence in *M. leprae* genomic clone MD0009-0-(B13) (354)
37	HinP#1-142	6,258	~69,000	~27,400	~741	nothing evident
38	HinP#2-4	6,567	~66,000	~24,400	~660	nothing evident
39	HinP#2-143	3,689	~65,000	~23,400	~633	Similar to sequence in *M. leprae* genomic clone B1529
40	HinP#2-145A	2,314	~64,000	~22,400	~606	nothing evident
41	HinP#2-147	7,021	65,000	23,400	~633	nothing evident
42	HinP#3-28	2,980	70,000	28,400	~768	Similar to *M. leprae* sequence in genomic clones MD0085 and sequence for *M. leprae* gp U00013 cds 27 of B1496 region
43	HinP#3-34	2,564	71,000	29,400	~795	Similar to sequence in *M. leprae* genomic clone B2168 (U00018 cds 9)
44	HinP#3-41	3,296	48,000	6,400	~1,728	Similar to antigen 85 complex protein subunit
45	HpaII#1-3	2,360	65,000	23,400	~633	Cytochrome C oxidase subunit II (156) Similar to *M. tuberculosis* sequence on clone MTCY 190.11c

TABLE 2

Immunostimulatory AP-fusion clones (cont'd)

No.	Clone Name	INF	Fus-MW	TBport	coding	Similarity (score)
46	HpaII#1-8	2,048	110,000	68,400	~1,848	nothing evident
47	HpaII#1-10	4,178	55,000	13,400	~633	Similar to immunogenic proteins MPB64/MPT64
48	HpaII#1-13	3,714	43,000	1,400	~39	nothing evident

Abbreviations: INF: pg/ml of INF-γ produced using fusion to stimulate immune T-cells. Fus. MW: Relative molecular weight of the fusion protein. TB port.: Estimated amount of fusion attributable to the *M. tuberculosis* protein. Coding: Amount of DNA needed to encode TB portion of fusion proteins. Similarity: Amino acid sequence similarity seen by analysis of DNA via the BLASTX or TBLASTX' programs. Scores for alignments are indicated in (). Due to the high G+C nature of M. TB DNA many false positives are evident. Only scores above 100 have good credibility.

2. DNA SEQUENCING AND DETERMINATION OF OPEN READING FRAMES

DNA sequence data for the sequences of the *Mycobacterium tuberculosis* DNA present in the clones shown in Tables 1 and 2 are shown in the accompanying Sequence Listing. The sequences are believed to represent the coding strand of the *Mycobacterium* DNA. In most instances, these sequences represent only partial sequences of the immunostimulatory peptides and, in turn, only partial sequences of *Mycobacterium tuberculosis* genes. However, each of the clones from which these sequences were derived encodes, by itself, at least one immunostimulatory T-cell epitope. As discussed in part V below, one of ordinary skill in the art will, given the information provided herein, readily be able to obtain the immunostimulatory peptides and corresponding full length *M. tuberculosis* genes using standard techniques. Accordingly, the nucleotide sequences of the present invention encompass not only those sequences presented in the sequence listings, but also the complete nucleotide sequence encoding the immunostimulatory peptides as well as the corresponding *M. tuberculosis* genes. The nucleotide abbreviations employed in the sequence listings are as follows in Table 3:

TABLE 3

Symbol	Meaning
A.....	A; adenine
C.....	C; cytosine
G.....	G; guanine
T.....	T; thymine
U.....	U; uracil
M.....	A or C
R.....	A or G
W.....	A or T/U
S.....	C or G
Y.....	C or T/U
K.....	G or T/U
V.....	A or C or G; not T/U
H.....	A or C or T/U; not G
D.....	A or G or T/U; not C
B.....	C or G or T/U; not A
N.....	(A or C or G or T/U) or (unknown or other or no base)
-.....	indeterminate*

* indicates an unreadable sequence compression.

The DNA sequences obtained were then analyzed with respect to the G+C content as a function of codon position over a window of 120 codons using the 'FRAME' computer program (Bibb, M.J.; Findlay, P.R.; and Johnson, M.W.; Gene 30: 157-166 (1984)). This program uses the bias of these nucleotides for each of the codon positions to enable the correct reading frame to be identified.

3. IDENTIFICATION OF T CELL EPITOPES IN THE IMMUNOSTIMULATORY PEPTIDES

The T-Site program, by Feller, D.C. and de la Cruz, V.F., MedImmune Inc., 19 Firstfield Rd., Gaithersburg, M.D. 20878, U.S.A., was used to predict T-cell epitopes from the determined coding sequences. It uses a series of four predictive algorithms. In particular, peptides were designed against regions indicated by the algorithm "A" motif which predicted alpha-helical periodicity (Margalit, H.; Spouge, J.L.; Cornette, J.L.; Cease, K.B.; DeLisi, C.; and Berzofsky, J.A., *J. Immunol.*, 138:2213 (1987)) and amphipathicity and those indicated by the algorithm "R" motif which identifies segments which display similarity to motifs known to be recognized by MHC class I and class II molecules (Rothbard, J.B. and Taylor, W.R., *EMBO J.* 7:93 (1988)). The other two algorithms identify classes of T-cell epitopes recognized in mice.

4. SYNTHESIS OF SYNTHETIC PEPTIDES CONTAINING T CELL EPITOPES IN IDENTIFIED IMMUNOSTIMULATORY PEPTIDES

A series of staggered peptides were designed to overlap regions indicated by the T-site analysis. These were synthesized by Chiron Mimotopes Pty. Ltd. (11055 Roselle St., San Diego, CA 92121, U.S.A.).

Peptides designed from sequences described in this application include:

Hin P#1-200 (6 peptides)

Peptide Sequence	Peptide Name
VHLATGMAETVASFSPS	HPI1-200/2
REVVHLATGMAETVASF	HPI1-200/3
RDSREVVHLATGMAETV	HPI1-200/4
DFNRDSREVVHLATGMA	HPI1-200/5
ISAAVVTGYLRWTTPDR	HPI1-200/6
AVVFLCAAAISAAVVTG	HPI1-200/7

AciI#2-827 (14 peptides)

Peptide Sequence	Peptide Name
VTDNPAWYRLTKFFGKL	CD-2/1/96/1
AWYRLTKFFGKLFLINF	CD-2/1/96/2
KFFGKLFLINFAIGVAT	CD-2/1/96/3
FLINFAIGVATGIVQEF	CD-2/1/96/4
AIGVATGIVQEFQFGMN	CD-2/1/96/5
TGIVQEFEFGMNWSEYS	CD-2/1/96/6
EFQFGMNWSEYSRFVGD	CD-2/1/96/7
MNWSEYSRFVGDVFGAP	CD-2/1/96/8
WSEYSRFVGDVFGAPLA	CD-2/1/96/9
EYSRFVGDVFGAPLAME	CD-2/1/96/10
SRFVGDVFGAPLAMESL	CD-2/1/96/11
WIFGWNRLPRLVHLACI	CD-2/1/96/12
WNRLPRLVHLACIWIVA	CD-2/1/96/13
GRAELSSIVVLLTNNTA	CD-2/1/96/14

HinP#1-3 (2 peptides)

Peptide Sequence	Peptide Name
GKTYDAYFTDAGGITPG	HPI1-3/2
YDAYFTDAGGITPGNSV	HPI1-3/3

HinP#1-3 / HinP#1-200 combined peptides

Peptide Sequences		Peptide Name
WPQGKTYDAYFTDAGGI	(HinP#1-3)	HPI1-3/1 (combined)
ATGMAETVASFSPSEGS	(HinP#1-200)	

AciI#2-823 (1 peptide)

Peptide Sequence	Peptide Name
GWERRLRHAVSPKDPAQ	AI2-823/1

HinP#1-31 (4 peptides)

Peptide Sequence	Peptide Name
TGSGETTTAAGTTASPG	HPI1-31/1
GAAILVAGLSGCSSNKS	HPI1-31/2
AVAGAAILVAGLSGCSS	HPI1-31/3
LTVAVAGAAILVAGLSG	HPI1-31/4

These synthetic peptides were resuspended in phosphate buffered saline to be tested to confirm their ability to function as T cell epitopes using the procedure described in part IV(B)(6) above.

5. CONFIRMATION OF IMMUNOSTIMULATORY CAPACITY USING T CELLS FROM TUBERCULOSIS PATIENTS

The synthetic peptides described above, along with a number of the PhoA fusion proteins shown to be immunostimulatory in mice were tested for their ability to stimulate gamma interferon production in T-cells from tuberculin positive people using the methods described in part IV(B)(6) above. For each assay, 5×10^5 mononuclear cells were stimulated with up to 1 μg/ml *M. tuberculosis* peptide or up to 50 ng/ml Phb A fusion protein. *M. tuberculosis* filtrate proteins, Con A and PHA were employed as positive controls. An assay was run with media alone to determine background levels, and Pho A protein was employed as a negative control.

The results, shown in Table 4 below, indicate that all of the peptides tested stimulated gamma interferon production from T-cells of a particular subject.

TABLE 4

Peptide or Pho A Fusion Protein Name	Concentration of Interferon-gamma (pg/ml)	Concentration of Interferon-gamma minus background (pg/ml)
CD-2/1/96/1	256.6	153.3
CD-2/1/96/9	187.6	84.3
CD-2/1/96/10	134.0	30.7
CD-2/1/96/11	141.6	38.3
CD-2/1/96/14	310.2	206.9
HPI1-3/2	136.3	23.0
HPI1-3/3	264.2	160.9
AciI 2-898	134.0	30.7
AciI 3-47	386.8	283.5
M. tuberculosis filtrate proteins (10 μg/ml)	256.6	153.3
M. tuberculosis filtrate proteins (5 μg/ml)	134.0	30.7
Con A (10 μg/ml)	2 839	2 735.7
PHA (1%)	10 378	10 274.7
Pho A control (10 μg/ml)	26.7	0
Background	103.3	0

V. CLONING OF FULL LENGTH *MYCOBACTERIUM TUBERCULOSIS* T-CELL EPITOPE ORFS

Most the sequences presented represent only part of a larger *M. tuberculosis* ORF. If desired, the full length *M. tuberculosis* ORFs that include these provided nucleotide sequences can be readily obtained by one of ordinary skill in the art, based on the sequence data provided herein.

A. GENERAL METHODOLOGIES

Methods for obtaining full length genes based on partial sequence information are standard in the art and are particularly simple for prokaryotic genomes. By way of example, the full length ORFs corresponding to the DNA sequences presented herein may be obtained by creating a library of *Mycobacterium tuberculosis* DNA in a plasmid, bacteriophage or phagemid vector and screening this library with a hybridization probe using standard colony hybridization techniques. The hybridization probe consists of an oligonucleotide derived from a DNA sequence according to the present invention labelled with a suitable marker to enable detection of hybridizing clones. Suitable markers include radionuclides, such as P-32 and non-radioactive markers, such as biotin. Methods for constructing suitable libraries, production and labelling of oligonucleotide probes and colony hybridization are standard laboratory procedures and are described in standard laboratory manuals, such as in reference nos. 15 and 16.

Having identified a clone that hybridizes with the oligonucleotide, the clone is identified and sequenced using standard methods such as described in Chapter 13 of reference no. 15. Determination of the translation initiation point of the DNA sequence enables the ORF to be located.

An alternative approach to cloning the full length ORFs corresponding to the DNA sequences provided herein is the use of the polymerase chain reaction (PCR). In particular, the inverse polymerase chain reaction (IPCR) is useful to isolate DNA sequences flanking a known sequence. Methods for amplification of flanking sequences by IPCR are described in Chapter 27 of reference no. 17 and in reference no. 23.

Accordingly, one aspect of the present invention is small oligonucleotides encompassed by the DNA sequences presented in the Sequence Listing. These small oligonucleotides are useful as hybridization probes and PCR primers that can be employed to clone the corresponding full length *Mycobacterium tuberculosis* ORFs. In preferred embodiments, these oligonucleotides will comprise at least 15 contiguous nucleotides of a DNA sequence set forth in the Sequence Listing, and in more preferred embodiments, such oligonucleotides will comprise at least 20 contiguous nucleotides of a DNA sequence set forth in the Sequence Listing.

One skilled in the art will appreciate that hybridization probes and PCR primers are not required to exactly match the target gene sequence to which they anneal. Therefore, in another embodiment, the oligonucleotides will comprise a sequence of at least 15 nucleotides and preferably at least 20 nucleotides, the oligonucleotide sequence being substantially similar to a DNA sequence set forth in the Sequence Listing. Preferably, such oligonucleotides will share at least about 75%-90% sequence identity with a DNA sequence set forth in the Sequence Listing and more preferably the shared sequence identity will be greater than 90%.

B. EXAMPLE - CLONING OF THE FULL LENGTH ORF CORRESPONDING TO CLONE HinP #2-92

Using the techniques described below, the full length gene corresponding to the clone HinP #2-92 was obtained. This gene, herein termed *mtb2-92* includes an open-reading frame of 1089 bp (identified based on the G+C content relating to codon position). The alternative 'GTG' start codon was used, and this was preceded (8 bps upstream) by a Shine-Dalgarno motif. The gene *mtb2-92* encoded a protein (termed MTB2-92) containing 363 amino acid residues with a predicted molecular weight of 40,436.4 Da.

Sequence homology comparisons of the predicted amino acid sequence of MTB2-92 with known proteins in the database indicated similarity to the cytochrome c oxidase subunit II of many different organisms. This integral membrane protein is part of the electron transport chain, subunits I and II forming the functional core of the enzyme complex.

1. CLONING THE FULL LENGTH GENE CORRESPONDING TO HinP #2-92

The plasmid pHin2-92 was restricted with either *Bam*H1 or *Eco*RI and then subcloned into the vector M13. The insert DNA fragments were sequenced under the direction of M13 universal sequencing primers (Yanisch-Perron, C. *et al.*, 1985) using the AmpliCycle thermal sequencing kit (Perkin Elmer, Applied Biosystems Division, 850 Lincoln Centre Drive, Foster City, CA 94404, U.S.A.). The 5'-partial MTB2-92 DNA sequence was aligned using a GeneWorks (Intelligenetics, Mountain View, CA, U.S.A.) program. Based on the sequence data obtained, two oligomers were synthesized. These oligonucleotides ($^{5'}$CCCAGCTTGTGATACAGGAGG$^{3'}$ $^{5'}$GGCCTCAGCGCGGCTCCGGAGG$^{3'}$) represented sequences upstream and downstream, over an 0.8 kb distance, of the sequence encoding the partial MTB2-92 protein in the alkaline phosphatase fusion.

A *Mycobacterium tuberculosis* genomic cosmid DNA library was screened using PCR (Sambrook, J. *et al.*, 1989) in order to obtain the full-length gene encoding the MTB2-92 protein. Two hundred and ninety-four bacterial colonies containing the cosmid library were pooled into 10 groups in 100 µl distilled water aliquots and boiled for 5 min. The samples were spun in a microfuge at maximal speed for 5 min. The supernatants were decanted and stored on ice prior to PCR analysis.

The 100 µl-PCR reaction contained: 10 µl supernatant containing cosmid DNA, 10 µl of 10X PCR-buffer, 250 µM dNTP's, 300 nM downstream and upstream primers, 1 unit *Taq* DNA polymerase.

The reactions were heated at 95°C for 2 min and then 40 cycles of DNA synthesis were performed (95°C for 30 s, 65°C for 1 min, 72°C for 2 min). The PCR products were loaded into a 1% agarose gel in TAE buffer (Sambrook, J. *et al.*, 1989) for analysis.

The supernatant, which produced 800 bp PCR products, was then further divided into 10 samples and the PCR reactions were performed again. The colony which had resulted in the correctly sized PCR product was then picked. The cosmid DNA from the positive clone (pG3) was prepared using the Wizard Mini-Prep Kit (Promega Corp, Madison, WI, U.S.A.). The cosmid DNA was further sequenced using specific oligonucleotide primers. The deduced amino acid sequence encoded by the MTB2-92 protein is shown in Fig. 1.

2. EXPRESSION OF THE FULL LENGTH GENE

To conveniently purify the recombinant protein, a histidine tag coding sequence was engineered immediately upstream of the start codon of *mtb2-92* using PCR. Two unique restriction enzyme sites for *Xba*I and *Hind*III were added to both ends of the PCR product for convenient subcloning. Two oligomers were used to direct the PCR reaction: (5'TCTAGACACCACCACCACCACCACGTGACACCTCGCGGGCCAGGTC3' and 5'AAGCTTCGCCATGCCGCCGGTAAGCGCC3')

The 100 µl PCR reaction contained: 1 µg pG3 template DNA, 250 µM dNTP's, 300 nM of each primer, 10 µl of 10X PCR buffer, 1 unit *Taq* DNA polymerase. The PCR DNA synthesis cycle was performed as above.

The 1.4 kb PCR products were purified and ligated into the cloning vector pGEM-T (Promega). Inserts were removed by digestion using both the *Xba*I and *Hind*III and the 1.4 kb fragment was directionally subcloned into the *Xba*I and *Hind*III sites of pMAL-c2 vector (New England Bio-Labs Ltd., 3397 American Drive, Unit 12, Mississauga, Ontario, L4V 1T8, Canada). The gene encoding MTB2-92 was fused, in frame, downstream of the maltose binding protein (MBP). This expression vector was named pMAL-MTB2-92.

3. PURIFICATION OF THE ENCODED PROTEIN

The plasmid pMAL-MTB2-92 was transformed into competent *E. coli* JM109 cells and a 1 litre culture was grown up in LB broth at 37°C to an OD_{550} of 0.5 to 0.6. The expression of the gene was induced by the addition of IPTG (0.5 mM) to the culture medium, after which the culture was grown for another 3 hours at 37°C with vigorous shaking. Cultures were spun in the centrifuge at 10,000 g for 30 min and the cell pellet was harvested. This was re-suspended in 50 ml of 20 mM Tris-HCl, pH 7.2, 200 mM NaCl, 1 mM EDTA supplemented with 10 mM β mercaptoethanol and stored at −20°C.

The frozen bacterial suspension was thawed in cold water (0°C), placed in an ice bath, and sonicated. The resulting cell lysate was then centrifuged at 10,000 g and 4°C for 30 min, the supernatant retained, diluted with 5 volumes of buffer A and applied to an amylose-resin column (New England Bio-Labs Ltd., 3397 American Drive, Unit 12, Mississauga, Ontario, L4V 1T8, Canada) which had been pre-equilibrated with buffer A. The column was then washed with buffer A until the eluate reached an A_{280} of 0.001 at which point, the bound MBP-MTB2-92 fusion protein was eluted with buffer A containing 10 mM maltose. The protein purified by the amylose-resin affinity column was about 84 kDa which corresponded to the expected size of the fusion protein (MBP: 42 kDa, MTB2-92 plus the histidine tag: 42 kDa).

The eluted MBP-MTB2-92 fusion protein was then cleaved with factor Xa to remove the MBP from the MTB2-92 protein. One ml of fusion protein (1 mg/ml) was mixed with 100 µl of factor Xa (200 µg/ml) and kept at room temperature overnight. The mixture was diluted with 10 ml of buffer B (5 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9, 6 M urea) and urea was added to the sample to a final concentration of 6 M urea. The sample was loaded onto the Ni-NTA column (QIAGEN, 9600 De Soto Ave., Chatsworth, CA 91311, U.S.A.) pre-equilibrated with buffer B. The column was washed with 10 volumes of buffer B and 6 volumes of buffer C (60 mM imidazole , 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9, 6 M urea). The bound protein was eluted with 6 volumes of buffer D (1 M imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9, 6 M urea).

At each stage of the protein purification, a sample was analysed by SDS polyacylamide gel electrophoresis (Laemmli, U.S. (1970) *Nature (London)*, 227:680-685) (see Fig. 2).

C. CORRECTION OF SEQUENCE ERRORS

It is noted that some of the sequences presented in the Sequence Listing contain sequence ambiguities. Naturally, in order to ensure that the immunostimulatory function is maintained, one would utilize a sequence without such ambiguities. For those sequences containing ambiguities, one would therefore utilize the sequence data provided in the Sequence Listing to design primers corresponding to each terminal of the provided sequence and, using these primers in conjunction with the polymerase chain reaction, synthesize the desired DNA molecule using *M. tuberculosis* genomic DNA as a template. Standard PCR methodologies, such as those described above, may be used to accomplish this.

VI. EXPRESSION AND PURIFICATION OF THE CLONED PEPTIDES

Having provided herein DNA sequences encoding *Mycobacterium tuberculosis* peptides having an immunostimulatory activity, as well as the corresponding full length *Mycobacterium tuberculosis* genes, one of skill in the art will be able to express and purify the peptides encoded by these sequences. Methods for expressing proteins by recombinant means in compatible prokaryotic or eukaryotic host cells are well known in the art and are discussed, for example, in reference nos. 15 and 16. Peptides expressed by the nucleotide sequences disclosed herein are useful for preparing vaccines effective against *M. tuberculosis* infection, for use in diagnostic assays and for raising antibodies that specifically recognize *M. tuberculosis* proteins. One method of purifying the peptides is that presented in part V(B) above.

The most commonly used prokaryotic host cells for expressing prokaryotic peptides are strains of Escherichia coli, although other prokaryotes, such as Bacillus subtilis *Streptomyces* or Pseudomonas may also be used, as is well known in the art. Partial or full-length DNA sequences, encoding an immunostimulatory peptide according to the present invention, may be ligated into bacterial expression vectors. One aspect of the present invention is thus a recombinant DNA vector including a nucleic acid molecule provided by the present invention. Another aspect is a transformed cell containing such a vector.

Methods for expressing large amounts of protein from a cloned gene introduced into *Escherichia coli* (*E. coli*) may be utilized for the purification of the *Mycobacterium tuberculosis* peptides. Methods and plasmid vectors for producing fusion proteins and intact native proteins in bacteria are described in reference no. 15 (ch. 17). Such fusion proteins may be made in large amounts, are relatively simple to purify, and can be used to produce antibodies. Native proteins can be produced in bacteria by placing a strong, regulated promoter and an efficient ribosome binding site upstream of the cloned gene. If low levels of protein are produced, additional steps may be taken to increase protein production; if high levels of protein are produced, purification is relatively easy.

Often, proteins expressed at high levels are found in insoluble inclusion bodies. Methods for extracting proteins from these aggregates are described in ch. 17 of reference no. 15. Vector systems suitable for the expression of *lacZ* fusion genes include the pUR series of vectors (24), pEX1-3 (25) and pMR100 (26). Vectors suitable for the production of intact native proteins include pKC30 (27), pKK177-3 (28) and pET-3 (29). Fusion proteins may be isolated from protein gels, lyophilized, ground into a powder and used as antigen preparations.

Mammalian or other eukaryotic host cells, such as those of yeast, filamentous fungi, plant, insect, amphibian or avian species, may also be used for protein expression, as is well known in the art. Examples of commonly used mammalian host cell lines are VERO and HeLa cells, Chinese hamster ovary (CHO) cells, and W138, BHK, and COS cell lines, although it will be appreciated by the skilled practitioner that other prokaryotic and eukaryotic cells and cell lines may be appropriate for a variety of purposes, e.g., to provide higher expression, desirable glycosylation patterns, or other features.

VII. SEQUENCE VARIANTS

It will be apparent to one skilled in the art that the immunostimulatory activity of the peptides encoded by the DNA sequences disclosed herein lies not in the precise nucleotide sequence of the DNA sequences, but rather in the epitopes inherent in the amino acid sequences encoded by the DNA sequences. It will therefore also be apparent that it is possible to recreate the immunostimulatory activity of one of these peptides by recreating the epitope, without necessarily recreating the exact DNA sequence. This could be achieved either by directly synthesizing the peptide (thereby circumventing the need to use the DNA sequences) or, alternatively, by designing a nucleic acid sequence that encodes for the epitope, but which differs, by reason of the redundancy of the genetic code, from the sequences disclosed herein.

Accordingly, the degeneracy of the genetic code further widens the scope of the present invention as it enables major variations in the nucleotide sequence of a DNA molecule while maintaining the amino acid sequence of the encoded protein. The genetic code and variations in nucleotide codons for particular amino acids is presented in Tables 5 and 6. Based upon the degeneracy of the genetic code, variant DNA molecules may be derived from the DNA sequences disclosed herein using standard DNA mutagenesis techniques, or by synthesis of DNA sequences.

TABLE 5

The Genetic Code

First Position (5' end)	Second Position				Third Position (3' end)
	T	C	A	G	
T	Phe	Ser	Tyr	Cys	T
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop (och)	Stop	A
	Leu	Ser	Stop (amb)	Trp	G
C	Leu	Pro	His	Arg	T
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	T
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	T
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val (Met)	Ala	Glu	Gly	G

"Stop (och)" stands for the ocre termination triplet, and "Stop (amb)" for the amber. ATG is the most common initiator codon; GTG usually codes for valine, but it can also code for methionine to initiate an mRNA chain.

TABLE 6

The Degeneracy of the Genetic Code

Number of Synonymous Codons	Amino Acid	Total Number of Codons
6	Leu, Ser, Arg	18
4	Gly, Pro, Ala, Val, Thr	20
3	Ile	3
2	Phe, Tyr, Cys, His, Gln, Glu, Asn, Asp, Lys	18
1	Met, Trp	2
Total number of codons for amino acids		61
Number of codons for termination		3
Total number of codons in genetic code		64

Additionally, standard mutagenesis techniques may be used to produce peptides which vary in amino acid sequence from the peptides encoded by the DNA molecules disclosed herein. However, such peptides will retain the essential characteristic of the peptides encoded by the DNA molecules disclosed herein, i.e. the ability to stimulate INF-γ production. This characteristic can readily be determined by the assay technique described above. Such variant peptides include those with variations in amino acid sequence including minor deletions, additions and substitutions.

While the site for introducing an amino acid sequence variation is predetermined, the mutation *per se* need not be predetermined. For example, in order to optimize the performance of a mutation at a given site, random mutagenesis may be conducted at the target codon or region and the expressed protein variants screened for the optimal combination of desired activity. Techniques for making substitution mutations at predetermined sites in DNA having a known sequence as described above are well known.

In order to maintain the functional epitope, preferred peptide variants will differ by only a small number of amino acids from the peptides encoded by the DNA sequences disclosed herein. Preferably, such variants will be amino acid substitutions of single residues. Substitutional variants are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the following Table 7 when it is desired to finely modulate the characteristics of the protein. Table 7 shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative substitutions. As noted, all such peptide variants are tested to confirm that they retain the ability to stimulate INF-γ production.

TABLE 7

Original Residue	Conservative Substitutions
Ala	ser
Arg	lys
Asn	gln, his
Asp	glu
Cys	ser
Gln	asn
Glu	asp
Gly	pro
His	asn; gln
Ile	leu, val
Leu	ile; val
Lys	arg; gln; glu
Met	leu; ile
Phe	met; leu; tyr
Ser	thr
Thr	ser
Trp	tyr
Tyr	trp; phe
Val	ile; leu

Substantial changes in immunological identity are made by selecting substitutions that are less conservative than those in Table 7, i.e., selecting residues that differ more significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in protein properties will be those in which (a) a hydrophilic residue, e.g., seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g., leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histadyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine. However, such variants must retain the ability to stimulate INF-γ production.

VIII. USE OF CLONED *MYCOBACTERIUM* SEQUENCES TO PRODUCE VACCINES

The purified peptides encoded by the nucleotide sequences of the present invention may be used directly as immunogens for vaccination. The conventional tuberculosis vaccine is the BCG (bacille Calmette-Guerin) vaccine, which is a live vaccine comprising attenuated *Mycobacterium bovis* bacteria. However, the use of this vaccine in a number of countries, including the U.S., has been limited because administration of the vaccine interferes with the use of the tuberculin skin test to detect infected individuals (see Cecil Textbook of Medicine (Ref. 33), pages 1733-1742 and section VIII (2) below).

The present invention provides a possible solution to the problems inherent in the use of the BCG vaccine in conjunction with the tuberculin skin test. The solution proposed is based upon the use of one or more of the immunostimulatory *M. tuberculosis* peptides disclosed herein as a vaccine and one or more different immunostimulatory *M. tuberculosis* peptides disclosed herein in the tuberculosis skin test (see section IX (2) below). If the immune system is primed with such a vaccine, it will be able to resist an infection by *M.*

tuberculosis. However, exposure to the vaccine peptides alone will not induce an immune response to those peptides that are reserved for use in the tuberculin skin test. Thus, the present invention would allow the clinician to distinguish between a vaccinated individual and an infected individual.

Methods for using purified peptides as vaccines are well known in the art and are described in the following publications: Pal and Horwitz (1992) (reference no. 8) (describing immunization with extra-cellular proteins of *Mycobacterium tuberculosis*); Yang et al. (1991) (reference no. 30) (vaccination with synthetic peptides corresponding to the amino acid sequence of a surface glycoprotein from *Leishmania major*); Andersen (1994) (reference no. 9) (vaccination using short-term culture filtrate containing proteins secreted by *Mycobacterium tuberculosis*); and Jardim et al. (1990) (reference no. 10) (vaccination with synthetic T-cell epitopes derived from *Leishmania* parasite). Methods for preparing vaccines which contain immunogenic peptide sequences are also disclosed in U.S. Patent Nos. 4,608,251, 4,601,903, 4,599,231, 4,5995230, 4,596,792 and 4,578,770. The formulation of peptide-based vaccines employing *M. tuberculosis* peptides is also discussed extensively in International Patent application WO 95/01441.

As is well known in the art, adjuvants such as Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA) may be used in formulations of purified peptides as vaccines. Accordingly, one embodiment of the present invention is a vaccine comprising one or more immunostimulatory *M. tuberculosis* peptides encoded by genes including a sequence shown in the attached sequence listing, together with a pharmaceutically acceptable adjuvant.

Additionally, the vaccines may be formulated using a peptide according to the present invention together with a pharmaceutically acceptable excipient such as water, saline, dextrose and glycerol. The vaccines may also include auxillary substances such as emulsifying agents and pH buffers.

It will be appreciated by one of skill in the art that vaccines formulated as described above may be administered in a number of ways including subcutaneous, intra-muscular and intra-venous injection. Doses of the vaccine administered will vary depending on the antigenicity of the particular peptide or peptide combination employed in the vaccine, and characteristics of the animal or human patient to be vaccinated. While the determination of individual doses will be within the skill of the administering physician, it is anticipated that doses of between 1 microgram and 1 milligram will be employed.

As with many vaccines, the vaccines of the present invention may routinely be administered several times over the course of a number of weeks to ensure that an effective immune response is triggered. As described in International Patent Application WO 95/01441, up to six doses of the vaccine may be administered over a course of several weeks, but more typically between one and four doses are administered. Where such multiple doses are administered, they will normally be administered at from two to twelve week intervals, more usually from three to five week intervals. Periodic boosters at intervals of 1-5 years, usually three years, will be desirable to maintain the desired levels of protective immunity.

As described in WO 95/01441, the course of the immunization may be followed by *in vitro* proliferation assays of PBL (peripheral blood lymphocytes) co-cultured with ESAT6 or ST-CF, and especially by measuring the levels of IFN-γ released from the primed lymphocytes. The assays are well known and are widely described in the literature, including in U.S. Patent Nos. 3,791,932; 4,174,384 and 3,949,064.

To ensure an effective immune response against tuberculosis infection, vaccines according to the present invention may be formulated with more than one immunostimulatory peptide encoded by the nucleotide sequences disclosed herein. In such cases, the amount of each purified peptide incorporated into the vaccine will be adjusted accordingly.

Alternatively, multiple immunostimulatory peptides may also be administered by expressing the nucleic acids encoding the peptides in a nonpathogenic microorganism, and using this transformed nonpathogenic

microorganism as a vaccine. As described in International Patent Application WO 95/01441, *Mycobacterium bovis* BCG may be employed for this purpose, although this approach would destroy the advantage outlined above to be gained from using separate classes of the peptides as vaccines and in the skin test. As disclosed in WO 95/01441, an immunostimulatory peptide of *M. tuberculosis* can be expressed in the BCG bacterium by transforming the BCG bacterium with a nucleotide sequence encoding the *M. tuberculosis* peptide. Thereafter, the BCG bacteria can be administered in the same manner as a conventional BCG vaccine. In particular embodiments, multiple copies of the *M. tuberculosis* sequence are transformed into the BCG bacteria to enhance the amount of *M. tuberculosis* peptide produced in the vaccine strain.

IX. USE OF CLONED *MYCOBACTERIUM* SEQUENCES IN DIAGNOSTIC ASSAYS

Another aspect of the present invention is a composition for diagnosing tuberculosis infection wherein the composition includes peptides encoded by the nucleotide sequences of the present invention. The invention also encompasses methods and compositions for detecting the presence of anti-tuberculosis antibodies, tuberculosis peptides and tuberculosis nucleic acid sequences in body samples. Three examples typify the various techniques that may be used to diagnose tuberculosis infection using the present invention: an in vitro ELISA assay, an in vivo skin test assay and a nucleic acid amplification assay.

A. IN VITRO ELISA ASSAY

One aspect of the invention is an ELISA that detects anti-tuberculosis mycobacterial antibodies in a medical specimen. An immunostimulatory peptide encoded by a nucleotide sequence of the present invention is employed as an antigen and is preferably bound to a solid matrix such as a crosslinked dextran such as SEPHADEX (Pharmacia, Piscataway, NJ), agarose, polystyrene, or the wells of a microtiter plate. The polypeptide is admixed with the specimen, such as human sputum, and the admixture is incubated for a sufficient time to allow antimycobacterial antibodies present in the sample to immunoreact with the polypeptide. The presence of the immunopositive immunoreaction is then determined using an ELISA assay.

In a preferred embodiment, the solid support to which the polypeptide is attached is the wall of a microtiter assay plate. After attachment of the polypeptide, any nonspecific binding sites on the microtiter well walls are blocked with a protein such as bovine serum albumin. Excess bovine serum albumin is removed by rinsing and the medical specimen is admixed with the polypeptide in the microtiter wells. After a sufficient incubation time, the microtiter wells are rinsed to remove excess sample and then a solution of a second antibody, capable of detecting human antibodies is added to the wells. This second antibody is typically linked to an enzyme such as peroxidase, alkaline phosphatase or glucose oxidase. For example, the second antibody may be a peroxidase-labeled goat anti-human antibody. After further incubation, excess amounts of the second antibody are removed by rinsing and a solution containing a substrate for the enzyme label (such as hydrogen peroxide for the peroxidase enzyme) and a color-forming dye precursor, such as o-phenylenediamine is added. The combination of mycobacterium peptide (bound to the wall of the well), the human antimycobacterial antibodies (from the specimen), the enzyme-conjugated anti-human antibody and the color substrate will produce a color than can be read using an instrument that determines optical density, such as a spectrophotometer. These readings can be compared to a control incubated with water in place of the human body sample, or, preferably, a human body sample known to be free of antimycobacterial antibodies. Positive readings indicate the presence of anti-mycobacterial antibodies in the specimen, which in turn indicate a prior exposure of the patient to tuberculosis.

B. SKIN TEST ASSAY

Alternatively, the presence of tuberculosis antibodies in a patient's body may be detected using an improved form of the tuberculin skin test, employing immunostimulatory peptides of the present invention. Conventionally, this test produces a positive result to one of the following conditions: the current presence of *M. tuberculosis* in the patient's body; past exposure of the patient to *M. tuberculosis*; and prior BCG vaccination. As

noted above, if one group of immunostimulatory peptides is reserved for use in vaccine preparations, and another group reserved for use in the improved skin test, then the skin test will not produce a positive response in individuals whose only exposure to tuberculosis antigens was via the vaccine. Accordingly, the improved skin test would be able to properly distinguish between infected individuals and vaccinated individuals.

The tuberculin skin test consists of an injection of proteins from *M. tuberculosis* that are injected intradermally. The test is described in detail in Cecil Textbook of Medicine (Ref. 33), pages 1733-1742. If the subject has reactive T-cells to the injected protein, the cells will migrate to the site of injection and cause a local inflammation. This inflammation, which is generally known as delayed type hypersensitivity (DTH) is indicative of *M. tuberculosis* antibodies in the patient's blood stream. Purified immunostimulatory peptides according to the present invention may be employed in the tuberculin skin test using the methods described in reference 33.

C. NUCLEIC ACID AMPLIFICATION

One aspect of the invention includes nucleic acid primers and probes derived from the sequences set forth in the attached sequence listing, as well as primers and probes derived from the full length genes that can be obtained using these sequences. These primers and probes can be used to detect the presence of *M. tuberculosis* nucleic acids in body samples and thus to diagnose infection. Methods for making primers and probes based on these sequences are well known and are described in section V above.

The detection of specific pathogen nucleic acid sequences in human body samples by polymerase chain reaction amplification (PCR) is discussed in detail in reference 17, in particular, part four of that reference. To detect *M. tuberculosis* sequences, primers based on the sequences disclosed herein would be synthesized, such that PCR amplification of a sample containing *M. tuberculosis* DNA would result in an amplified fragment of a predicted size. If necessary, the presence of this fragment following amplification of the sample nucleic acid could be detected by dot blot analysis (see chapter 48 of reference 17). PCR amplification employing primers based on the sequences disclosed herein may also be employed to quantify the amounts of *M. tuberculosis* nucleic acid present in a particular sample (see chapters 8 and 9 of reference 17). Reverse-transcription PCR using these primers may also be utilized to detect the presence of *M. tuberculosis* RNA, indicative of an active infection.

Alternatively, probes based on the nucleic acid sequences described herein may be labelled with suitable labels (such a P^{32} or biotin) and used in hybridization assays to detect the presence of *M. tuberculosis* nucleic acid in provided samples.

X. USE OF CLONED *MYCOBACTERIUM* SEQUENCES TO RAISE ANTIBODIES

Monoclonal antibodies may be produced to the purified *M. tuberculosis* peptides for diagnostic purposes. Substantially pure *M. tuberculosis* peptide suitable for use as an immunogen is isolated from the transfected or transformed cells as described above. The concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few milligrams per milliliter. Monoclonal antibody to the protein can then be prepared as follows:

A. MONOCLONAL ANTIBODY PRODUCTION BY HYBRIDOMA FUSION.

Monoclonal antibody to epitopes of the *M. tuberculosis* peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler and Milstein (1975) or derivative methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected purified protein over a period of a few weeks. The mouse is then sacrificed, and the antibody-producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall (1980), and derivative

methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Harlow and Lane (1988).

B. ANTIBODIES RAISED AGAINST SYNTHETIC PEPTIDES.

An alternative approach to raising antibodies against the *M. tuberculosis* peptides is to use synthetic peptides synthesized on a commercially available peptide synthesizer based upon the amino acid sequence of the peptides predicted from nucleotide sequence data.

In a preferred embodiment of the present invention, monoclonal antibodies that recognize a specific *M. tuberculosis* peptide are produced. Optimally, monoclonal antibodies will be specific to each peptide, i.e. such antibodies recognize and bind one *M. tuberculosis* peptide and do not substantially recognize or bind to other proteins, including those found in healthy human cells.

The determination that an antibody specifically detects a particular *M. tuberculosis* peptide is made by any one of a number of standard immunoassay methods; for instance, the Western blotting technique (Sambrook et al., 1989). To determine that a given antibody preparation (such as one produced in a mouse) specifically detects one *M. tuberculosis* peptide by Western blotting, total cellular protein is extracted from a sample of human sputum from a healthy patient and from sputum from a patient suffering from tuberculosis. As a positive control, total cellular protein is also extracted from *M. tuberculosis* cells grown in vitro. These protein preparations are then electrophoresed on a sodium dodecyl sulfate-polyacrylamide gel. Thereafter, the proteins are transferred to a membrane (for example, nitrocellulose) by Western blotting, and the antibody preparation is incubated with the membrane. After washing the membrane to remove non-specifically bound antibodies, the presence of specifically bound antibodies is detected by the use of an anti-mouse antibody conjugated to an enzyme such as alkaline phosphatase; application of the substrate 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium results in the production of a dense blue compound by immuno-localized alkaline phosphatase. Antibodies which specifically detect the *M. tuberculosis* protein will, by this technique, be shown to bind to the *M. tuberculosis*-extracted sample at a particular protein band (which will be localized at a given position on the gel determined by its molecular weight) and to the proteins extracted from the sputum from the tuberculosis patient. No significant binding will be detected to proteins from the healthy patient sputum. Non-specific binding of the antibody to other proteins may occur and may be detectable as a weak signal on the Western blot. The non-specific nature of this binding will be recognized by one skilled in the art by the weak signal obtained on the Western blot relative to the strong primary signal arising from the specific antibody-tuberculosis protein binding. Preferably, no antibody would be found to bind to proteins extracted from healthy donor sputum.

Antibodies that specifically recognize a *M. tuberculosis* peptide encoded by the nucleotide sequences disclosed herein are useful in diagnosing the presence of tuberculosis antigens in patients.

All publications and published patent documents cited in this specification are incorporated herein by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

XI. REFERENCES

1. Skamene, E. (1989). Genetic control of susceptibility to Mycobacterial infections. Ref. Infect. Dis. 11:S394-S399.

2. Kaufmann, S.H.E. (1991). Role of T-Cell Subsets in Bacterial Infections. Current Opinion in Immunology 3:465-470.

3. Orme, I.M., et al. (1992). T Lymphocytes Mediating Protection and Cellular Cytolysis During the Course of Mycobacterium-Tuberculosis Infection - Evidence for Different Kinetics and Recognition of a Wide Spectrum of Protein Antigens. Journal of Immunology 148:189-196.

4. Daugelat, S., et al. (1992). Secreted Antigens of Mycobacterium tuberculosis: characterization with T Lymphocytes from Patients and Contacts after Two-Dimensional Separation. J. Infect. Dis. 166:186-190.

5. Barnes et al. (1989). Characterization of T Cell Antigens Associated with the Cell Wall Protein-Peptidoglycan Complex of *Mycobacterium tuberculosis*. J. Immunol. 143:2656-2662.

6. Collins et al. (1988). Biological activity of protein antigens isolated from *Mycobacterium tuberculosis* culture filtrate. Infect. Immun. 56:1260-1266.

7. Lamb et al. (1989). Identification of Mycobacterial Antigens Recognized by T Lymphocytes. Rev. Infect. Dis. 11:S443-S447.

8. Pal, P.G., et al. (1992). Immunization with Extracellular Proteins of Mycobacterium tuberculosis Induces Cell-Mediated Immune Responses and Substantial Protective Immunity in a Guinea Pig Model of Pulmonary Tuberculosis. Infect. Immun. 60:4781-4792.

9. Andersen (1994). Infection & Immunity 62:2536.

10. Jardim et al. (1990). Immunoprotective *Leishmania major* Synthetic T Cell Epitopes. J. Exp. Med. 172:645-648.

11. Orme et al. (1993). Cytokine Secretion by CD4 T Lymphocytes Acquired in Response to *Mycobacterium tuberculosis* Infection. J. Immunology 151:518-525.

12. Boesen et al. (1995). Human T-Cell Responses to Secreted Antigen Fractions of *Mycobacterium tuberculosis*. Infection and Immunity 63:1491-1497.

13. Mougneau et al. (1995). Expression Cloning of a Protective *Leishmania* Antigen. Science 268:536-566.

14. Yang et al. (1990). Oral *Salmonella typhimurium* (AroA$^-$) Vaccine Expressing a Major Leishmanial Surface Protein (gp63) Preferentially Induces T Helper 1 Cells and Protective Immunity Against Leishmaniasis. J. Immunology 145:2281-2285.

15. Sambrook et al. (1989). Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, ed. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.

16. Ausubel et al., (1987). Current Protocols in Molecular Biology, ed. Greene Publishing and Wiley-Interscience: New York (with periodic updates).

17. Innis et al., (1990). PCR Protocols: A Guide to Methods and Applications, Academic Press: San Diego.

18. Kanehisa (1984). Nuc. Acids Res. 12:203-213, 1984.

19. Wetmur et al. (1968). J. Mol. Biol. 31:349-370.

20. Beaucage et al. (1981) Tetra. Letts. 22:1859-1862.

21. Matteucci et al. (1981). J. Am. Chem. Soc. 103:3185.

22. Jacobs et al. (1991) METHODS IN ENZYMOLOGY 204:537-555.

23. Earp et al. (1990). Nucleic Acids Research 18:3721-3729.

24. Ruther et al. (1983). EMBO J. 2:1791.

25. Stanley and Luzio (1984). EMBO J. 3:1429.

26. Gray et al. (1982). Proc. Natl. Acad. Sci. USA 79:6598.

27. Shimatake and Rosenberg (1981). Nature 292:128.

28. Amann and Brosius (1985). Gene 40:183.

29. Studiar and Moffatt (1986). J. Mol. Biol. 189:113.

30. Yang et al. (1991). Identification and Characterization of Host-Protective T-Cell Epitopes of a Major Surface Glycoprotein (gp63) from *Leishmania major*. Immunology 72:3-9.

31. Mdluli et al. (1995). New vectors for the in vitro generation of alkaline phosphatase fusions to proteins encoded by G+C-rich DNA. Gene 155:133-134.

32. Lim et al. (1995). Identification of *Mycobacterium tuberculosis* DNA Sequences Encoding Exported Proteins by Using *phoA* Gene Fusions. J. Bact. 177:59-65).

33. Cecil Textbook of Medicine, (1992, 19th edition), Wyngaarden et al, eds. W.B. Saunders, Philadelphia, PA.

34. Hubbard et al. (1992). Immunization of mice with mycobacterial culture filtrate culture proteins. Clin. exp. Immunol. 87: 94-98.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANTS: UNIVERSITY OF VICTORIA

(ii) TITLE OF INVENTION: *MYCOBACTERIUM TUBERCULOSIS* DNA SEQUENCES ENCODING IMMUNOSTIMULATORY PEPTIDES

(iii) NUMBER OF SEQUENCES: 76

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Klarquist Sparkman Campbell Leigh & Whinston, LLP

(B) STREET: One World Trade Center, Suite 1600, 121 S.W. Salmon Street

(C) CITY: Portland

(D) STATE: OR

(E) COUNTRY: USA

(F) ZIP: 97204-2988

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Disk, 3.5-inch

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: MS DOS

(D) SOFTWARE: WordPerfect 5.1+

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 06/000,254

(B) FILING DATE: 06/15/95

(viii) ATTORNEY/AGENT INFORMATION

(A) NAME: Richard J. Polley

(B) REGISTRATION NUMBER: 28,107

(C) REFERENCE/DOCKET NUMBER: 2847-45176/RJP

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (503) 226-7391

(B) TELEFAX: (503) 228-9446

(2) INFORMATION FOR SEQ ID NO: 1

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 265

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#1-62

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 1

ACGCGGACCT CGAAGTTCAT CATCGAGTGA TACGTGCCAC ACATCTCGGC 50

GCAGTGGCCC ACGAATGCAN CCGGTCTTGG TGATTTCNTC GATCTGGAAG 100

ACGTTGACCG ARTTGTTTGC CACCGGGTTA GGCATCACGT CACGCTTGAA 150

CAAGAACTCC GGCACCCAGA ATGCGTGTGT CACATCGGCT GAGGCCATTT 200

GGAATTCGAT ACGCTTGCCG GACGGCAGCA CCAGCACCGG AATTTCGGTG 250

CTGTGCAACG TCTCG 265

(2) INFORMATION FOR SEQ ID NO: 2

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 484

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#1-152

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 2

CTGGTACGAC GCCGGCAAGG ACTACGGACG AGGTGGCACA GAATTCAATG 50

CGGCGCTCAT CGGAACCGAC GTGCCCGACG NCGTTTGCTC GACGACGATG 100

GTGNTTCCAN TTCGCCTNAN CGGTGTNCTG ACTGCCNTTG ACGACCTGNT 150

CGGCCARGTT GGGNTGGACA CAACGGATTA CGTCGATTCG CTGCTGGCCG 200

ACTATGAGTT CAACGGCCGC CATTACGCTG TGCCGTATGC TCGCTCGACG 250

CCGCTGTTCT ACTACAACAA GGCGGCGTGG CAACAGGCCG GCCTACCCGA 300

CCGCGGACCG CAATCCTGGT CAGAGTTCGA CGAGTGGGGT CCGGAGTTAC 350

AGCGCGTGGT CGNCGCCGGT CGATCGGCGC ACGGCTGCGT AACGCCGACC 400

TCATCTCGTG GACGTTTCAG GGACCGAACT GGGCATNCGG CGGTGCCTAC 450
TCCGACAAGT GGACATTGAC ATTGACCGAG CCCG 484

(2) INFORMATION FOR SEQ ID NO: 3

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 513

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#1-239

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 3

GGCGGCCAGA CGTCGGAACT CGCGGCCAAT TGGTGTGGTG GGAACCGCGA 50
TCCTCGACGC AACCGCTTCG CGGTCTTGGC AGTGTTCGAT GCCAATCTGC 100
CGGCCGGGAC GCTGCCGGAT GCGGCCCGTT CACCGAGGCT GGTGACAAGA 150
CCTGGCGTTG TCGTTCCGGG CACTACTCCC NAGGTCGGTC AAGGCACCGT 200
CAAAGTGTTC AGGTATACCG TCGAGATCGA GAACGGTCTT GATCCCACAA 250
TGTACGGCGG TGACAANNNN ATTCGCCCAG ATGGTCGACC AGACGTTGAC 300
CAATCCCAAG GGCTGGACCC ACAATCCGCA ATTCGGCGTT CGTGCGGATC 350
GACAGCGGAA AACCCGACTT CCGGATTTCG CTGGTGTCGC CGACGACAGT 400
GCGCGGGGGN TGTGGCTACG AATTCCGGCT CGAGACGTCC TGCTACAACC 450
CGTCGTTCGG CGGCATGGAT CGCCAATCGC GGGTGTTCAT CAACGAGGCG 500
CGCTGGGTAC GCG 513

(2) INFORMATION FOR SEQ ID NO: 4

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 510

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#1-247

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 4

GTGTGCAACC AGTGTGTGTN CGTGTGCGAA CCAGTGTGTA GTGGTAACCA 50
GGACCACGTT GCAAACCAGT GTTGGAGTGC AGTGTTGCGT GCNAGTGTTG 100
CNCGTTGCAG TGTTNGNCGA GCCGAGATTG GAAGTTNCCG ACATTACCGT 150
TGCCGACGTT GCCCTCGCCG ACGTTCGCCA AGCCCAGGTT GCGGACACGC 200
CGGTGATTGT GCGTGGGGCA ATGACGGGCT GCTGGCCCGG CCGAATTCCA 250
AGGCGTCGAT CGGCACGGTG TTCCAGGACC GGGCCGCTCG CTACGGTGAC 300
CGAGTCTTCC TGAAATTCGG CGATCAGCAG CTGACCTACC GCGACCGTAA 350
CGCCACCGCC AACCGGTNNG CCGCGGTGTT GGCCNNNCGC GGCGTCGGCC 400
CCGGCGACGT CGTTGGCATC ATGTTGCGTA ACTCACCCAG CACAGTCTTG 450
GCGATGCTGG CCACGGTCAA GTGCGGCGTA TCGCCGGCAT GCTCAACTAC 500
CACCAGCGCG 510

(2) INFORMATION FOR SEQ ID NO: 5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 456

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#1-426

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 5

GCAACGGAGA GGTGGACTAT GCCGGACCGG CACCGCGAAG GGGTTGGTGC 50
CGGCCCGGGT GGTGACGGTG CACATTCTGC GCAATTCGCT GAGTTCCGGT 100
GGTGACCTTC CTGGGCGCGG AGTCTGGGCG CGCTGATGGC GGAGCGAKTG 150
TGACCGAAGG AANTCNGTTC AACATCCACG GCGTCGGGGG CGTGCTGTAT 200
CAAGCGGTCA CCGTCAGGAG ACGCCGACGG TGGTGTCGAT CGTGACGGTG 250
CTGGTGCTGA TCTACCTGAT CACCAATCTG TTGGTGGATC TGCTGTATGC 300
GGCCCTGGAC GCCGNNGATN CGCTATGGCT GAGCACACGG GGTTCTGGCT 350
CGATGCCTNG CGCGGGTTGC GCCGGCGTCC TAAANTCGTG ATCGCGCGGC 400
GCTGAKCCTG CTGATTCTTG TCGTGGCGGC GTTTCCGTCG TTGTTTACCG 450
CAGCCG 456

(2) INFORMATION FOR SEQ ID NO: 6

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 175

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 6

TCNCTTANYC CTTCANCTGN CATCTNTCCC AANNACCGAA NTCTGGACCT 50

ATSACGNCCA NCTNAANATG NCCCNCGACN AAGGNCNTTG NACGTTCNCT 100

GKACCACCAN CGGGTTGCAT SCCAAGCTAG NCGAACATCA NASGTTNCGC 150

GCNTACGAGC CGACCCGCCG CGGCG 175

(2) INFORMATION FOR SEQ ID NO: 7

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 231

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-23

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 7

CTTCTCGCGC CAGCCGTCCC GCTGTCCGGG ATGCGCTACC GGTCGTCAGC 50

GCCAAGACGG TGCAGCTCAA CGACGGCGGG TTGGTGCGCA CGGTGCACTT 100

GCCGGCCCCC AATGTSGCGG GGCTGCTGAG TGCGGCCGCG TGCCGCTGTT 150

GCAAANNGCG ACCACGTGGT GCCCGCCGCG ACGGCCCCGA TCGTCGAAGG 200

CATGCAGATC CAGGTGACCC GCAAATCGGA T 231

(2) INFORMATION FOR SEQ ID NO: 8

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 173

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-26

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 8

GTTCGNCGCG CTCAAAAGGT TGACGATGGT CACGTCGCAC GTGCTGGCCG 50
AGACCAAGGT GGATTTCGGT GAAGACCTCA AAGANCTCTA CTCGNATCGT 100
CAAGGCCCTC AACGACGACC GAAAGGATTT CGTCACCTCG CTGCAGCTGT 150
TGCTGACGTT CCCATTTCCC AAC 173

(2) INFORMATION FOR SEQ ID NO: 9

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 223

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 9

CCTGTTNCAA CGGTNCNTTC NCGGAACGGA CGACTTCTGA TNCGNNCTCG 50
GNCGTTCCCT CGCACCGGTC GATGGTGATC AAGGTCAGCG TCTTCGCGGT 100
GGTCATGCTG CTGGTGGCCG CCGGTCTGGT GGTGGTATTC GGGGACTTCC 150
GGTTTGGTCC CACAACCGTC TACCACGCCA CCTTCACCGA CNCGTNGCGG 200
CTGAANGCAG GCCAGAAGGT TCG 223

(2) INFORMATION FOR SEQ ID NO: 10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 120

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-272

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 10

CAACGAGATC GCACCCGTGA TTAGGAGGTG ACGGTGGCAG CGCCGACCCC 50
GTCGAATCGG ATCGAAGTAA CGCTCCGTAG ACGCCAGCTC GTCCGCGCCG 100
ATGCCGACCT GCCACCCGTG 120

(2) INFORMATION FOR SEQ ID NO: 11

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 160

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-506

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 11

CNGGCNNCCA NCGGGTGCGC CAWGCACGGC CGGTCCGTGC GAGATCGTCN 50

CNAATGGCAN GCCGGCGCCC AAKANANNNC CGGTACCGTG CCTTCGTNGW 100

GCAWCCTNGC GACCAACCCC GAGATYGCYA CNCTACNGCC GGKACATGAC 150

CGTGGTGCGG 160

(2) INFORMATION FOR SEQ ID NO: 12

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 133

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-508

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 12

GACTGGNCCC GAYGYTGTGN CCGGHNCGTH GGNCGHGCHG CANTCGAYCC 50

TGGCCGTTGC TTCGGTGCCG GGTTGTTCAT CGCCTTCGAC CAGTTGTGGC 100

GCTGGAACAG CATAGTGGCG CTAGTGCTAT CGG 133

(2) INFORMATION FOR SEQ ID NO: 13

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 421

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#2-511
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 13

GCGNACNCTG CGCATNGCTG CCNGTANCCC GGCGCCNAGG CATGAGNCNN 50
TAGGCCGAAA TGCCTGGTKA ANCTNGCGTG TSGTGGTTGA CCCGCNGCGT 100
SCNGGCNTAC AKGTGCATGC TGTNGATCGG CAGTGGGAGA GGTGAGCGGT 150
GCGGCGTNAA GGTGCGGAGG TTNGASNTCT GGCGGTGTCG GCGTTNGGTG 200
GCTTTGTTCC CGGCGGTCGC GGGGTGCTCC NGNATTCCGG CGACNAACNA 250
AANNCCGGGN AGSACGAYNC CCGTCGACAC CNGGCAAACG CTGAGGGCCG 300
GCACGGACCC TTCTTCCCGC AATGTGGCGG CGTCAGCGAT CANGACGGTG 350
ACCGAGCTGW ACAAGGGTGA CCGGGCTGGT CAACACCGCC AAGAAGTCGG 400
TGGGCTNCCA ATGGCNTGGC G 421

(2) INFORMATION FOR SEQ ID NO: 14
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#2-523
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 14

CCAGNCCNCC NAACNTGTYN CGNTCTCAYY TCGCCGTCGC TGCCGGTNCG 50
TGTGTGCACC ATCTGCACCG ACCCGTGKAA CYTCGATCAC GANACTGGNA 100
GAGNTCAGGC ATNAAAGCCG GAGTGGCACA GCAACGGTCG CTACTGGAAT 150
TGGCGAAGCT GGATGCTGAG CTGAC 175

(2) INFORMATION FOR SEQ ID NO: 15
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 263
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-639

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 15

GGGCTGGATT CGAGGCTCGT GCATGNCGTA CGACTANGGG TAGCGCCCAG 50

CTGCTCAATA CCATCGGTTG GATAACAAAG GCTGAACATG AATGGCNTGA 100

TCTCNACAAG CGTGCGGCTC CCACCGACCC CGGCGCCCCT CGAGCCTGGG 150

GSTGTCGCGA TCCTGATCGC GGCGACACTT TTCGCGACTG TCGTTGCGGG 200

GTGCGGGAAA AAACCGACCA CGGCGAGCTC CCGAGTCCCG GGTCGCCGTC 250

GCCGGAAGCC CAC 263

(2) INFORMATION FOR SEQ ID NO: 16

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 168

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-822

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 16

YGCCATGCGA AGCGCACCCC GGTCCGGAAG NCCTGCACAG TTCWNCCGTG 50

CTCGCCGCGA CGCTACTCCT CGNYTGCGGC GGTCCCAYGC AGCCAYGCAG 100

CATCACCTTG ACCTTTATCC GCAACGYGYA ATYCCAGGCC AAYGCCGAYG 150

GGATCATCGA YACCKACA 168

(2) INFORMATION FOR SEQ ID NO: 17

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 181

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-854

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 17

ACCNGTTCCC GCCGGNCTNA CNCNCGGTGC CGTTGCACCG GCCANCTGCA 50
GCCTGCCCCG ACGCCGAAGT GGTGTTCGCN CCGCGGCCGC TTCGAACCGC 100
CCGGGATTGG CACGGTCGGC AABGCATTCG TCAGCNNTGC GCTCGAAGGT 150
CAACAAGAAT GTCGGGGTCT ACGCGGTGAA A 181

(2) INFORMATION FOR SEQ ID NO: 18

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 95

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-872

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 18

AGGTKACGGT GGCAGCGCCG ACCCCGTCGA ATCGGWTCGA AGAAYGCTCC 50
GKACACGCCA GCTGCGTCCG YGCCGATGCC GACCTGCCAC CCGTG 95

(2) INFORMATION FOR SEQ ID NO: 19

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 65

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-884d

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 19

AKCGGTCACC KACGGGCCGG CCACCGATGC GATTGTCAAC GGATTCCAAG 50
TGGTTGYGCA TGCGC 65

(2) INFORMATION FOR SEQ ID NO: 20

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 156

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-8841

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 20 --

TCTTCTACAA GGACGCCTTC GCCAAGCACC AGGAGCTGTT CGACGACTTG 50

GNCGTCAACG TCAACAATGG CTTGTCCGAT CTGTACRAGC AAGWTCGAGT 100

CGCTGCCGNB CGCAACGCGA CGAGATCATC GAGGACCTAC ACCGTTGCCA 150

CGAACA 156

(2) INFORMATION FOR SEQ ID NO: 21

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 123

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-8941

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 21

ATNCCGTTCC ACTNCCGCGG CAGCAGCTGG NTTTGCGCAC ACGGTGACCC 50

AGTGGCGNTT GGTGGGGCCT CGCTGACGGC GAGTNTGGNC GAGCGTCCTC 100

GGTCGGTGNC CTNTCNTCCC GCC 123

(2) INFORMATION FOR SEQ ID NO: 22

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 636

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-898

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 22

CGGTCWHKCA ANTTGATGBC NGCGCGCAAG GCCGNCATGG TNGAGATGCC 50
AACCACACCA CCGGCTGGNT CCGCATGGAC TTCGTGNTTS CCAGTCGCNG 100
CCTGATTGGG TGNCGCACCG ACNNCCTNCA CCGAGACCSG TGGCTC-SGA 150
GGANCTCGAC ATCAATKCAN CCGGAGNAGN ANGCTGACCN AACATNCGCT 200
CATCGACCGC GGATGTCNAT CGAGNACGST GCCAAGSCGC TGCAGCTGGA 250
TNCTCGAGCG CGCCATGGAG TNATRTGCGS CCGACGAATN CGTCGAGGTG 300
ACCCCGGAGA NTCGTGCGGA TSCGCRAAGT CGAGCTGGCC GGCCNGCCGC 350
CCGGGCTNMG CAGCCGGGCG CGCACCNAAG GCGCGTGGCN TAGCANACTT 400
GGCGNGCTGG CCGCGCGAGC GTANACNGCC ACTGCGAAAN TCCANGCCCG 450
GCTTTTCGCA GCCGGGTTNA CGCTCGTGGG GGTACTGGAT AGCCTGATGG 500
GCGTGCCCAG NCCCANGTCC GCCGCGTCTG TGTGACGGTC GGCGCGTTGG 550
TCGCGCTGGC GTGTATGGTG TTGGCCGGGT GCACGGTCAG CCCGCCGCCG 600
GCACCCCAGA GCASTGATAC GCCGCGCAGC ACACCG 636

(2) INFORMATION FOR SEQ ID NO: 23

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 103

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-916

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 23

CTTCCGGCGG GACAACAACA GGTCTCACCG GCGCCACACC CTGACACCTG 50
ATCGCGTCTG CCGATCCCGG TCGGAGCACC CGGGTTCCAC CGCTGTGCCC 100
CCC 103

(2) INFORMATION FOR SEQ ID NO: 24

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 207

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-1014

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 24

GCCACCGGTT CATCGCGTGG TGCTGGTCAC CGCCNGGAAN GCCTCAGCGG 50
ATCCCCTGCT GCCACCGCCG CCTATCCCTG CCCCAGTCTC GGCGCCGGCA 100
ACAGTCCCGY CCGTGCAGAA CCTCACGGCT NCTHCCGGGC GGGAGCAGCA 150
ACAGGTTCTC ACCGGYGCCW NGYACCCGCA CCGATCGCGT CGCCGATTCC 200
GGTCGGA 207

(2) INFORMATION FOR SEQ ID NO: 25

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 204

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-1025

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 25

TTNCGCANNC GTTCATCCAG GTCCACTGGT GTCGCANCTC TCNNTGATGC 50
ACCGGTTCCG GATATATGTC NACATCNCCS TCSTCGTCCT GGTGCTGGTA 100
CTNACGAACC TGATCGCGCA TTTCACCACA CCGTGNGCGA GCATCGCCAC 150
CGTCCCGGCC GCCYGCGGTC GGACTGGTGA TCTTGGTKCG GAGTAGAGGC 200
CTGG 204

(2) INFORMATION FOR SEQ ID NO: 26

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 207

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-1035

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 26

ATACCNGTCA TCCNGCACAT NGTCAACCTN GAGTCGGTNC TCACCTACGA 50
GGCACGCCCG AGATGCATCA CTGGTGCTCG RTCAGNCCTT CACGGCTTGG 100

CCGCCTTCCG GTAGGACCGT HGCATGCCCG TCTTCGGCGC CTCGGGTGTT 150
CGGTCCTGGC TCTCGGGCTG CTGGCCNCTG CGCCCCACCC CGCACCGGGC 200
CGGCTTC 207

(2) INFORMATION FOR SEQ ID NO: 27

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-1084

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 27

YCNAGNCKCG TNATNGCSGN CKCATNTNAC NGGANCCNGG ATTNCSTACG 50
CCACNGTGAT CGCGCTGGTN GCCGCGCTGG TGGCGCGTGT ACGTGCTCTC 100
GTCCACCGGN AANTAAGCGC ACCATCGTGG GCTACTTCAC CTCTGCTGTC 150
GGGCTCTATC CCGGTGACCA GGTCCGCGTC CTGGGCGTCC NGGTGGGTGA 200
GATCGACATG ATCGAGCCGC GGTCGTCCGA CGTSAAGATC ACTATGTCGG 250
TGTCCAAGGA CGTCAAGGTG CCCGTGSACG NTGCAGGCC 289

(2) INFORMATION FOR SEQ ID NO: 28

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 198

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-1089

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 28

TTGNACCANG CCTATCGCAA GCCAATCACC TATGACACGC TGTGGCAGGC 50
TGACACCGAT CCGCTGCCAG TCGTCTTCCC CATTGTGCAA GGTGAACTGA 100
GCAANGCAGA CCGGACAACA GGTATCGATA GCGCCGAATG CCGGCTTGGA 150
CCCGGTGAAT TATCAGAACT TYGCAGTCAC GAACGACGGG GTGATTTT 198

(2) INFORMATION FOR SEQ ID NO: 29

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 149
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#2-1090
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 29

TCACGANGGT RYNACMGCAA CWCGACCGCC ACGTCASGCC GCCGCGCACG 50
AAGATCACCG TGCCTGCNCG ATGGGTCGTG AACGGAATAG AAYGCAGCGG 100
TGAGGTCAAN YGCGAAGCCG GGAACCAAAT CCGGTGACCG CGTCGGCAT 149

(2) INFORMATION FOR SEQ ID NO: 30
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#2-1104
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 30

GGACCCGCCA AGCATCAGCC GGTCAACAGC CGCCGCCGGT GGCCAAAGTT 50
CGAGCAGCCG CCGGTATCGT GCTCGGCCCG GCTAGACCAA AAACTTTACG 100
CCAGCGCCCG AAGCCACCCG ACTCCAAGGC CTCGGCCCGG TTGGGTTCGC 150
ACATGGGTGA GTTCTATATG CCCTACCCGG GCACCCGGTT CAACCAGGAA 200
ACCGTCTCGC 210

(2) INFORMATION FOR SEQ ID NO: 31
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 255
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#3-9
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 31

CAGNCCGCTG NCCCGGAACT GTTCCAGCAG CTACAAGACC TTCGACAACG 50
TNGCGCGTCA ACCTGCANTC GAGCGCAACC TCTCGGTGGC GCTCAACGAG 100
TGTTCGCCGG CTTCAACCCG CTGGACCCGC GAAACCTCGA CGTGTCCCCG 150
CTGCCTTCGC TGGCCAAGCG CGCCGCCGAC ATCCTGCGCC AGGACGTGGG 200
CGGGCAGGTC GACATTTTCG ATGTCAATGT GCCCACCATC CAGTACGACC 250
AGAGC 255

(2) INFORMATION FOR SEQ ID NO: 31
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 164
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#3-12
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 31

AAYNCCNGGC CRTCGACGGT NCCGGTTCNC RCCACCGGTC TATATCCACC 50
CGGGTCNRCA TTMANANTGA NTMNCCGCCG GTGCGGCCGT CGAGCGTGAC 100
CTGGCATCCC CTGAGACGCT GCTGGGTTGC CCCGGGGAGN TCGAMANTCG 150
GGCATCGCAC CATC 164

(2) INFORMATION FOR SEQ ID NO: 32
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 237
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#3-15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 32

ACGGACGGCA ACGGGATGCG ACCCGATCCC ACCGGTCGCC ACGAGGGACG 50
CTACTTCGTC GCCGGGCAGC CGANCCGACC GTCNGTTCNG CGANGGCGAC 100
NGCCGAAGCC GTTGACCCAC NTTGGTCAGC AGCAGCTGGA TSAGTCAGGT 150
GCCGTTGGTG TTTCGCCGTC AGCGGTGTCG GGGTGGGTGC GTTCTGGGCA 200
CCGTCGACTG TGGTGGGCGC TNGCGGGCGN TGGTGGC 237

(2) INFORMATION FOR SEQ ID NO: 33

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 374

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#3-47

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 33

CNGATNGCTC GGNCTNCGGT ACCNAACTCG NAACTCGCGC CCWYGCGNAC 50
GCAGGNCCGC GGTTGGCACC ACCAGCGACA TCAATCANGC AGGWKNCCCG 100
CCACGTTGCA AGACGGCGGC AATCTTCGCC TGTCGCTCAC CGACTTTCCG 150
CCCAACTTCA ACATCTTGCA CATCGACGGC AACAACGCCG AGGTCGCGGC 200
GATGATGAAA GCCACCTTGC CGCGCGCGTT CATCATCGGA CCGGACGGCT 250
CGNACGNACG GTCGACACCA ACTACTTCAC CAGCATCGAG CTGACCAGGA 300
CCGCCCCGCA GGTGGTCACC TACACCATCA ATCCCGAGGC GGTGTGGTCC 350
GACGGGACCC CGATCACCTG GCCG 374

(2) INFORMATION FOR SEQ ID NO: 34

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#3-78 (overlaps with AciI#3-167)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 34

GAGAACTCCG GGCCGANTTT TGGACA 26

(2) INFORMATION FOR SEQ ID NO: 35

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 204

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#3-133

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 35

TGTCGGGTNA RNGTTCGCGT CCATGATTGC TCTTGCAACG CTGTTGACGC 50

TTATCAATCA AGTCGTCGGC ACTCCGTATA TTCCCGGTGG CGATTCTCCC 100

GCCGGGACCG ACTGCTCGGA GCTGGCTTCG TGGGTATCGA ATGCGGCGAC 150

GGCCAGGCCG GTTTTCGGAG ATAGGTTCAA CACCGGCAAC GAGGAAGCGC 200

CTTG 204

(2) INFORMATION FOR SEQ ID NO: 36

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 312

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#3-134

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 36

CANNTTAGAC TGTCGTGACA TATCNCNNTN TACNCNTGGN ACGGCCATNA 50

TTGGATAATN CGTGATAANC ACCACAAGAA TNATTCCTAT GNATATTGTC 100

GGTACGTTCG CGNCCATGAT TNGCTCTTGC AACGCTGTTG ACGCTTATCA 150

ATCAAGTCGT CGNCACTCCG TATATTCCCG GTGNCGATTC TCCCGCCGGG 200

ACCGACTGCT CRGAGCTGGC TTCGTGGGTA TCGAATGCGS CGACGSCCAG 250

GCCGGTTTTC GSAGATAGGT TCAACACCGG CAACGAGGAA GCGCCTTGGC 300

GGCTCGGGGC TN 312

(2) INFORMATION FOR SEQ ID NO: 37

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 676

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#3-166

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 37

AGGCCAATCG NTGATGCGAC TCGAACGGGT TCGGCGCCGA TGACTGTTTC 50

GCGAAGTTCA TCAGCACCCT CGTTGGCGCG AAGGGCACGA CGGTGTACCG 100

GWWRYSAMKA CRCYGCYATG AGTYTCTGCS TGTATTGCGG TGCSGAGCTT 150

GCCGACCCGA CCAGGTGCGG KGCGTGNCGG CSCAKACWAG ATTGGTTCAA 200

CCTGGCNATC GGACCNACGA CGCCGACGGT CGGCGCCGCG ACGACGGCAN 250

ACGGNATNGC GACCCGANTC CNYACCNGGT CGCCACGAGG GACGNCTACT 300

TCGTCGCCNG GCAGCCGACC GANCTCGTTN NNCGCGASGN CGACGCCGAA 350

GCCGTTGACC CACTTGGTCA GCAGCAGCTG GNNATCANGN TCANGGTGCC 400

GTTNNGGTGT TTCGCCGTCA GCGGTGTCGG GGTGGGTGCG TTCTGGGCAC 450

CGTCGACTGT GGTGGGCGCT TGCGGGCGTG GTGGCGTTTC TCGGGCTGGT 500

GGGAGCCGGT GTCGTCGGGA CGCTGTTCCT GAATCGAGAC CGGGAGTCCA 550

TCGACGACAA GTACCTCGCN CCTTGAGGCG GTCCGGACTC ACCGGTGAGT 600

TCAACTCCGA CGCGAACGCC ATCGCCCGCS GCAAGCAGGT GTGCCGCCAG 650

TTGCANASAC GGTGGCGAAC AGCNSA 676

(2) INFORMATION FOR SEQ ID NO: 38

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 853

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#3-167

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 38

GTGNGCGCGC CNTCGAGCAN GTCTTGGCNG CGANCCCGAB ACAANTGATT 50
CCCGACATCC GGTACACACC GAACCCCNAA NCGATGCGCC NGGCGGCCCG 100
CTGGTAGAAA GGGGAAATCG CCAGTGCTGA CTCGCKTCAT CCGACGCCAG 150
TTGAKCCKTT TKGCGAKCGT CKCCGTAGTG GCAATCGTCG TATTGGGCTG 200
GTACTACCTG CGAATTCCGA GTCTGGTGGG TNGTCGSGCA GTACACCTTG 250
AAGGCCGACT TGCCCGNATC GGGTGGCCTG TATCCGACGG CCAATGTGAC 300
CTACCGCGGT ATCACCATTG GCAAGGTTAC TGCCGTCGAG SCCACCGACC 350
AGGGCNGCAC GANGTGACGA TGAGCATCGC CAGNCAACTA SAAAATCSCC 400
GTCGATGCCT NCGGCGAACG TGCATTCGGN GTCAGCGGTN GGCGAGCAGT 450
ACATCGACCT NGTGTCCACC GGTGCTCCGG GTNAAATACT TCTCCTCCGG 500
ACAGACCATC ACCAANGGCA CCGTTCCCAG TGAGATCGGG CCGGCGCTGG 550
ACAANTCCSA ATCNGCGGGT TGGCCGCATT NGCCCACGGA GAAGATCGGC 600
TTGCTGCTCG ACGAGACNGC GCAAGCGGTG GGTGGGCTGG GACCCGCGNN 650
TTGCAACGGT TGGTCGATTC CACTCAAGCG ATCGTCGGTG ACTTCAAAAC 700
CAACATTGGC GACGTCAACG ACATCATCGA GAACTCCGGG CCGATTTTGG 750
ACAGCCAGGT CAACACGGGT GATCAGATCG ACGCTGGGCG CGCAAATTGA 800
ACAATSTGGC CGCACAGACC GCNGACCAGG GAKCAGAACG TGCGAAGCAT 850
CCT 853

(2) INFORMATION FOR SEQ ID NO: 39

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 209

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#3-204

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 39

GCGGTTGGCA CCACCAGCGA -AATCAGCAG GNDCCCGCCA CGTTGCAAGA 50
CGGCGGCAAT CTTCGCCTGT CGCTCACCGA CTTTCCGCCC AACTTCAACA 100
TCTTGCACAT CGACGGCAAB AABGCCGAGG TCGCGGCGAT GATGAAAGCC 150
ACCTTGCCGC GCGCGTTCAT CATCGGACCG GACGGCTCGA CGACGGTCGA 200
CACCAACTA 209

(2) INFORMATION FOR SEQ ID NO: 40

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 166

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#3-206
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 40

AGATCGTCAG TGAGCAGAAC CCCGCCAAAC CGGCCGCCCG AGGTGTTGTT 50
CSAGGGCTGA AGNCNCTGCT CGCGACGGTC GCTGCTGGCC GTCGTCGGGA 100
TCGGGCTTGG CTCGCGCTGT ACTTCACGCC GGCGATGTCG NCCCGCGAGA 150
TCGTGTATCA TCGGGT 166

(2) INFORMATION FOR SEQ ID NO: 41
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 221
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#3-214
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 42

CCAGNTCCTC NNATATCGAC ACCCTCNACN AAGACCGCTT CGCGAGATCA 50
ACNCTCAGAT ATNCNNACTA TCNCCNNTNC ACGCACACCT CAACATNANA 100
NAATNGAACT ATNGNCTTCG CCTCACCACC AAGGTTCAGG TTANCGGCTG 150
NCGTTTKCTC TKCGCCGGCT CGAACACGCC ATCGTGCGCC GGKACACCCG 200
GATGTTTGAC GACCCGCTGC A 221

(2) INFORMATION FOR SEQ ID NO: 43
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 117
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#3-281

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 43

CGGYCCGNNC AAYYYGNCGC GCHNCGGYGY AGAGGTCGNY AAGGTCGCCA 50
AGGTAACGCT GATCGAYGGG NACANGCAAG TATTGGTGNA CTTCACCGTG 100
GHTHGCTHGC TGTYAGC 117

(2) INFORMATION FOR SEQ ID NO: 44

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 385

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: BsaHI#1-21

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 44

GAACCTCCTC GCCCGCGCTT GGCCTAGCAT TAATCGACTG GCACGACAGT 50
TGCCCGACTG GGTACACGGC ATGGACGCAA CGCGAATGAA TGTGAGTTAG 100
CTCACTCATT AGGCACCCCA GGCGTTGACA CTTTATGCTT CCGGCTCGTG 150
TAGTTGTGTG GGAATTGTGG AGCGGATAAC AATTTCGACG ACGAGGAAAC 200
AGCTGTAGAC ATGGATTGAC GAATTTGAAT ACGACTCACT ATAGGAATTC 250
GAGCTCGGTA CCCGGGGATC CTCTAGAGTC CTTCGCCGCG GGTCGCCACC 300
ATCAGGGCCA GTGCGATCGC AAGCGCGGGG TACCGGGCGC CATAGTCTTC 350
AGCATCGGCG TGTTGACCGC AGAGACCGGA CGGGG 385

(2) INFORMATION FOR SEQ ID NO: 45

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 285

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#1-12

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 45

CCCGCAGCAG TACCCGCAGN CCCACACCCG CTATNCGCAG CCCGAACAGT 50
TCGGTGCACA GCCCACCCNA GCTCGGCGTG CCCGGTCAGT ACGGCCAATA 100
CCAGCAGCCG GGCCAATATG NCCAGCCGGN ACAGTNACGN CCAGCCCGGC 150
CAGTACGCNA CCGCCCGGTC AGTACCCCGG GCAATACGGC CCGTATGNCC 200
AGTCGGGTCA GGGGTCGAAG CGTTCGGTTG CGGTGATCGG CGGCGTGATC 250
GCCGTGATGG CCGTGCTGTT CATCGGCGCG GTTCT 285

(2) INFORMATION FOR SEQ ID NO: 46

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 186

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#1-142

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 46

GCNCGTGNCC GTGCCGCCCG GTTGAACGTG AGCNGCTGNC NATNGCCCCA 50
GCCGAGACGA GAACGTCCCC GAGGAGTATG CAGACTGGGA AGACGCCGAA 100
GACTATGACG ACTATGACGA CTATGAGGCC GCAGACCAGG AGGCCGCACG 150
GTCGGCATCC TGGCGACGGC GGTTGCGGGT NCGGTT 186

(2) INFORMATION FOR SEQ ID NO: 47

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 402

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#1-144

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 47

GTCGCTGAAT GTGTTGTCGG AGACCGTGAT CAGACCTATC CGCACCTGAG 50
CGCCGCCTCC ACGGGTGGCT AAGTTCTCCG ACACCATCGG CAAGCGCGAC 100
GAGCAGACTC ANGCACCTAC TAGCCCAGGC CAACCAGGTG GCCAGCATCC 150

TGGGTGATCG CAGTGAGCAG GTCGACCGCC TATTGGTCAA CGCTAAGACC 200
CTGATCGCCG CGTTNCAACR GASNGCGCCG CGCGGTCGAC GCCCTGCTGG 250
GGAACATCTC CGCTTTCTCG CCCAGGYGCA AAACCTTCAT SAACGACAAN 300
CCGAACCTGA ACCATGTGCT CGAGCNGCGC ATCCTSACSA CCTGTTGGTS 350
GACSGCAAGG AGGATTTGGC TGAAANCCTN ACGATSTTGG GCAGAKTCAG 400
CG 402

(2) INFORMATION FOR SEQ ID NO: 48

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 468

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#1-200

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 48

AGNCCGTGCA CTGGAANCTT CGGCTCAGWT GTCTCCGATG TGGACGGCAA 50
SGCTGATGAT CTCCCGGTTG GAAGTCGANT CGATKASAAA TGGCTTGGCG 100
GCTGGTGGTG TTCGATGCCT GGCACCRACT GGCBACGATC NSCGCCTGGN 150
CGCGATCGGC GCTTAGCTCG GCTGGNNCCC TGTGGTGGGT TTCGACGTGC 200
TCGGTGTTGG TGCTGCTGGT GGTCGAAGGT GTGGCAATCA ACGTTCTGGC 250
TGTTGCGTCG TGATTCGGTA ACCGTCGGTA CCGACGACGA TGCGCCCGGG 300
CTGCGACTGG CCGTTGTCTT CCTGTGCNNG CCGCCGCGAT CTCGGCGGCN 350
GTGGTGACTG GGTACCTGCG CTGGACGACA CCGGACCGCG ACTTCAATCG 400
GGATTCCCGG GAAGTGGTGC ATCTTGCCAC GGGGATGGCC GAGACGGTCG 450
CGTCATTCTC CCCGAGCG 468

(2) INFORMATION FOR SEQ ID NO: 49

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 417

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#2-23

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 49

GTCCAAGGCC GTAGCCCACC TCCTGGAAGT CGTACCACGT CGACTCGACC 50
AGGACGGCTG CAGTCAGCAC TTCGTCAACC CGCGATCATC AACGTGCACC 100
TACGGCAGTG TGACGCACCC CGGACCATCG CACTGGCCGG GGTTCACACG 150
CCGAACACTG CTGACCGCAC TGGATCTGCT GGTCGCATGC ACCACTTCAA 200
GGTGGTGACG TACCTCAAAA TGGGTTTCCC GTTGTCCACC GAGGAAGTCC 250
CGCTGATTCA TGGGCAATAA CGCTCCCTAT CCGCAGTGTC ACCAGTGGGT 300
GCAAGCGGCG ATGGCCAAGT TGGTCGCTGA CCACCCCGAC TACGTTTTCA 350
CAACCTCGAC TCGACCGTGG AACATCAAAC CCGGCGATGT GATGCCAGCA 400
ACCTATGTCG GGATCTG 417

(2) INFORMATION FOR SEQ ID NO: 50

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 279

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#2-143

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 50

CGGTCGAGCC GATGAACGTC TGCAGTTCAC CGCAACCACG CTCAGCGGTG 50
CTCCCTTCGA TGCGCAAGCC TGCAAGGCAA TGCCGCGGTG TTGTGGTTCT 100
GGACGCCGTG GTGCCCGTTC TGCAACTGTC AGAAGCCCCC AGCCGCAGCC 150
AGGTAGCGGC CGCTAATCCG GCGGTCACCT TCGTCGGAAT CGCCACCCGC 200
GCCGACGTCG GGGCGATGCA GAGCTTTGTC TCGAAGTACA ACCTGAATTT 250
CACCAACCTC AATGACGCCG ATGGTGTGA 279

(2) INFORMATION FOR SEQ ID NO: 51

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 324

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#2-145

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 51

CGGCCCGGCG GCGCCCTGGT GAAGCTTGGA GAATGGGTGA GCGCAGCTGC 50

CCACCACACG GGACCGGTGC GGACGCGSTG ACGCGCCTGG TGGTCAGCAN 100

CNTGGCCGGT CTGCTGTTGT ATGCCAGCTT CCCGCCGCGC AACTGCTGGT 150

GGCGGCGGTG GTTGGGCTNC GCATTGCTGG CCTGGGTGCT GACCCACCGC 200

GCGACGACAC CGGTGGGTGG GCTGGGCTAC GGCCTGCTAT TCGGCCTGGT 250

GTTCTACGTC TCGTTGTTGC CGTGGATCGG CGAGCTGGTG CNCCGGGCCC 300

TGGTTGGCAC TGNCGACGAC GTGC 324

(2) INFORMATION FOR SEQ ID NO: 52

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 229

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#2-150

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 52

CCAGGCTAGC ACGTATGCTC CGGCTCGTTG TGTGTGGAAT GTGAGCGGAT 50

GACANKNCAC ACAGGADAYA GCTATGACNA TGATTACGCC AAGCTATTTA 100

GGTGABACTA TAGAATAYTC AAGCTATGCA TCCAAYGCGT TGGGAGCTCT 150

YCCATATGGT CGACCTGCAY GCGGCCGCAC TAGTGATTST THGCGCCGGC 200

NYGCWGCGGC NYAYGACCGC YAAYACCAC 229

(2) INFORMATION FOR SEQ ID NO: 53

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 293

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#3-28

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 53

CCACACAACA CAAATCTACG TCGTAATGCA GTCGTAAGTC CATCCGACGT 50
CGATGGCAAG GACAGCACCC GACGGCCAAC GGCATATACA TCGTCGGCTC 100
GCCGGTCACA AGCACATCAT CATGGACTCG TCCACTACGG CGTACCCGTC 150
AACTCGCCCA ACGGATATCG CACCGATGTC GACTGGCCAC CCAGATCTCC 200
TACAGCGGTG TCTTCGTGCA CTCAGCGCCG TGGTCGGTGG GGGCTCAGGG 250
CCACACCAAC ACCAGCCATG GCTGCCTGAA CGTCAGCCCG AGC 293

(2) INFORMATION FOR SEQ ID NO: 54

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 816
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#3-30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 54

CGNCGYCGSC GNGCSCTAYC GGTGCGGGAG GGTACAYCCA AGCANTCCGG 50
GACCGGCCGT CYCGCYGGGA ACGCCGTGCT CCTACAYACC GGCGRCGGGC 100
GCGTTGCCAC GSCCCGACAC CCCACTACCC NGNCGCGGGC GCCACCRTTG 150
GCCCGTTNMG GTGGACCCGA NCTTCCCGGC ACCGCTCGAT GTCCAGCCGT 200
CGCCGCCTAA TCCCGATGGG CCGCMGCCGA CKCCGGGCAT CCTAAGTGCT 250
GGGCGGCCGG GCGAGCCGGN TCCGGNTGTT CCGGCATACC GWTGCCSYTG 300
CCGNCGAAC- TGCACGCACC CAACCGCTTG AGCCGTTTCC TGACGGGACG 350
GGAGGTAGCA ACCAATGAGC ACCATCTTCG AYATCCGSAG CCTGCKACTN 400
GYCGAWACTG TCTNGCAAAG GTAGTGGTCG TCGGCGGGTT GGTGGTGGTC 450
TTGGCGGTCG TRGCCGNCTG NCRGCCGGCG CGCRGCTCTA CCGGAAACTG 500
ACTANACTAC CGTGGTCGCR TATTTTCTST GAGGCGCTCG CGCTGTACCC 550
AGGAGASAAA GTCCAGATCA TGGGTGTGCG GGTCGGTTCT ATCGACAAGA 600
TCGAGCCGGC CGGCGACAAG ATGCGAGTCA CGTTGCACTA NCAGCAASAA 650
ATACCAGGTG CCGGCCACGC TACCGNYGNW CGMTCCTCAA CCCCAGCCTG 700
GTGGCCTCGC GCACCATCCA GCTGTCACCN NCGTACACCG GCGGCCCGGT 750
CTTGCAAGAC GGCGCGGTGA TSCCAATCGA GCGCACCCAG RTGCCCGTCG 800
AGTGGGATCA GTTGCG 816

(2) INFORMATION FOR SEQ ID NO: 55

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 117
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: HinPI#3-34
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 55

CAGCCACCTC GTTCGCCGCC GACATCGACT ATCAGCCGAC CCGGCCACTG 50
CTGACCTGAT CGCCAACAGC TGGAGGCCCT ACCGGCTGCA GTTCAATTCA 100
CCCGCTGCGG GTCGGCG 117

(2) INFORMATION FOR SEQ ID NO: 56
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 242
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: HinPI#3-41
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 56

AGGTGTCGTG CTTCATGCCT GGCGCCCAAT CCAGTTTCTA CACCGACTGG 50
TATCACCCTT CGCAGACAAA CGGCCAGAAC TACACCTACA AGTGGGAGAC 100
CTTCCTTACC ACACAGATGC CCGCCTGGCT ACAGGCCAAC AAGGCGTGTC 150
CCCCACAGGC AACGCGGCGG TGGGTCTTTC GATCTCGGGC GGTTCCGCGC 200
TGACCCTGGC CGCGTACTAC CCGCAGCAGT TCCCGTACGC CG 242

(2) INFORMATION FOR SEQ ID NO: 57
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 340
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HpaII#1-3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 57

TGCTGCAGAT AGCCAAGGAT CCAGTCGTGA TTGATATCAC GTCTTTCCAG 50

TGAATTGAAG TTTGGCTATC AAAGGGTGAA CTTSAAAGAC GGCACACTGA 100

CCTATGATGG TGCCGATCCG GAGCGCAAGC GCGCCATGGT TTCCAAGCCA 150

GAGGGCAAGN ACAAGTACGG CGAAGAGCTG GTCGGGCCGG TGCGCGGGCT 200

CAACACCGAG GACCGGACCT ACCTGAATTT CGACAAGGTC GAGACGTTGG 250

GCAGCAGCAC CGAAATTCCG GTGCTGGTGC TGCCGTCCGG CAAGCGTATC 300

GAATTCCAAA TGGCCTCAGC CGATGTGATA CACGCATTCT 340

(2) INFORMATION FOR SEQ ID NO: 58

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HpaII#1-8

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 58

CNGACTCCAA CNAGTGCGNT CAANCNGNTG TNCCNGACAA GAAGGTTCCT 50

ACATCCGCAA NTCGGTGNAA NGCCACTGTG GATGCCTACG ACGGAACGGT 100

CACGCTGTAC CAACAGGACG NAAAAGGATC CGGTGCTCAA GGCCTGGATG 150

CAGGTCTTCC CCGGCACGGT AAAGCCTAAG AGCGACATTG CGCCGGAGCT 200

TGCCGAGCAN CTGCGGTATC CCGAGGACCT GTTCAAGGTG CAGCGCATGT 250

TGTTGGCCAA AT 262

(2) INFORMATION FOR SEQ ID NO: 59

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 241

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HpaII#1-10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 59

CCACCANNNA ACRRCACAGC TCCGGCCRRC CGTNCGCAGG CCACCCGCAN 50

CGTAGTGCTC AAATTCTTCC AGGACCTCGG TGGGGYACAT CCGTCCACCT 100

GGTACAAGGC CTTCAACTAC AACCTCGCGA CCTCGCAGCC CATCACCTTC 150

GACACGTTGT TCGTGCCCGG CACCACGCCA CTGGACAGCA TCTACCCCAT 200

CGTTCAGCGC GAGCTGGCAC GTCAGACCGG TTTCGGTGCC G 241

(2) INFORMATION FOR SEQ ID NO: 60

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 243

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HpaII#1-13

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 60

CCGGCGGATC TGCGTGACGA NTGTATNCCA CGGNACTACC CGCGGTCCTT 50

CCTCNANTNC CGCCGGNCCA GNCGCAGNCT NCNGATGTCC NGCTATAACC 100

TGCGCGATCG CCGCCGGGCT GCCCGACAAC ACGGTGNGCG CCGCCGCTGC 150

TTCCGCCAAT TCTGGGTGNC GGCATNCCGG CAGCGCCCGG CCCAGCACTG 200

AGAGGGGGAC GTTGATGCGG TGGCCGACGG CGTGGCTGCT GGC 243

(2) INFORMATION FOR SEQ ID NO: 61

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2346

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-825

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 61

GCGCTGNCAT TCGNACTTCG GACNGCGTTN GCGGTGGTGC TGATCATGAA 50
NCTACGACGG CGCCACCGGC AGCTTCCCGT CATGGGTGCT CTATCCCTGT 100
GCGCTGGCCA TGATGGTGTT CTCGAATKCG TTCAGCGTNC TGCGCAGCGC 150
AGTGANACCG AGGGTGATGC CGCCAACCAT CGACTTGGTC CGGGTCAACT 200
CACGGCTGAC CGTGTTCGGC CTGCTCGGCG GCACCATCGC TGGTGGCGCG 250
ATTGCGGCCG GAGTCGAATT CGTCTGCACC CACCTGTTCC AGCTGCCGGG 300
CGCGTTGTTC GTCGTCGTCG CGATCACCAT CNNTNNNGCT TCGCTGTCGA 350
TNCNCATTCC GCGCTGGGTC GAGGTGACCA GCGGTGAGGT CCCGGCCACA 400
TTGAGCTACC ACCGGGATAG GGNCAGACTA CGGCGACNGC TGGCCGGAGG 450
AAGTCAAGAA CCTCGGCGGA ACACTCCGAC AACCGTTGGG CCGCAACATC 500
ATTACCTCCC TGTGGGGTAA CTGCACCATC AAGGTGATGG TCGGCTTTCT 550
GTTCTTGTAT CCGGCGTTTG TCGCCAAGGC GCACGAAGCC AACGGGTGGG 600
TGCAATTGGG CATGCTGGGC CTGATCGGCG CGGCGGCCGC GGTCGGCAAC 650
TTCGCCGGCA ATTTCACCAG CGCACGCCTG CAGCTAGGCA GGCCAGCTGT 700
GCKGGTNGTG CGCTGCACCG TGCTAGTTAC CGTGTTAGCC ATCGCGGCCG 750
CGGTGGCCGG CAGCCTGGCA GCGACAGCNA TTGCCACCCT GATCACGGCA 800
GGGTCCAGTG CCATTGCTAA AGCCTCGCTG GACGCCTCGT TGCAGCACGA 850
CCTGCCCGAG GAGTCGCGGG CATCGGGGTT TGGGCGTTCC GAGTCGACTC 900
TTCAGCTGGC CTGGGTGCTG GGCGGCGCGG TGGGCGTGTT GGTGTACACC 950
GAGCTGTGGG TGGGCTTCAC TGCGGTGAGC GCGCTGCTGA TCCTGGGTCT 1000
GGCTCAGACC ATCGTCAGCT TCCGCGGCGA TTCGCTGATC CCTGGCCTGG 1050
GCGGTAATCG GCCCGTGATG GCCGAGCAAG AAACCACCCG TCGTGGTGCG 1100
GCGGTGGCGC CGNAGTGAAG CGCGGTGTCG CAACGCTGCC GGTGATCCTG 1150
GTGATTCTGC TCTCGGTGGC GGCCGGGGCC GGTGCATGGC TGCTAGTACG 1200
CGGACACGGT CCGCAGCAAC CCGAGATCAG CGCTTACTCG CACGGGCACC 1250
TGACCCGCGT GGGGCCCTAT TTGTACTGCA ACGTGGTCGA CCTCGACGAC 1300
TGTCAGACCC CGCANGCGCA GGGCGAATTG CCGGTAAGCG AACGCTATCC 1350
CGTGCAGCTC TCGGTACCCG AAGTCATTTC CCGGGCGCCG TGGCGTTTGC 1400
TGCAGGTATA CCAGGACCCC GCCAACACCA CCAGCACCTT GTTTCGGCCG 1450
GACACCCGGT TGGCGGTCAC CATCCCCACT GTCGACCCGC AGCGCGGGCG 1500
GCTGACCGGG ATTGTCGTGC AGTTGCTGAC GTTGGTGGTC GACCACTCGG 1550
GTGAACTACG CGACGTNCGC ACGCGGAATG GTCGGTGCGC CTTATCTTTT 1600
GACGAGGCCG CGGCTCGACG NC-CCTTAAG CGCGGTCGGC GCCAACGGTC 1650
CGAAGAGCCG CCGACACCCG GGGCACATCG GCGCATCATG GAACTGTGCG 1700
GATCGGAGTC GGGGTTTGCA CCACGCCCGA CGCGCGGCAG GCCGCGGTGG 1750
AGGCTGCGGG CCAGGCGCGC GACGAGCTGG CGGGTGAGGC GCCGTCGCTG 1800
GCGGTGTTGC TTGGATCGCG TGCACACACC GACCGGGCTG CCGACGTCCT 1850
GAGCGCGGTG CTGCAGATGA TCGACCCGCC CGCGCTTGTC GGTTGCATCG 1900

CCCAGGCCAT CGTCGCCGGC CGCCACGAGA TCGAGGACGA GCCCGCGGTG 1950
GTGGTGTGGC TGGCGTCCGG CTTGGCCGCC GAGACATTCC AGCTGGACTT 2000
TGTCNGTACC GGCTCGGGTG CCCTGATCAC CGGTTATCGG TTCGACCGNA 2050
CCGCCCGGGA TCTGCATCTG CTGCTGCCGG ACCCGTACAC ATTCCCGTCG 2100
AACCTGCTCA TCGAGCACCC CAACACCGAC CTGCCGGGCA CCGCNGTCGT 2150
GGGCGGCGNT GGTGAGCGGC GGGCGCCGGC GGGGCGACAC CCGGSTGTKC 2200
CGCGATCACG ACGTGCTCAC CTCCGGMGTC GTCGGCGTGC GCCTGCSCGG 2250
GATGCGCGGT GTMCCGGTCG TGTCGCAGGG TTGNCGGCCG ATCGGCTACC 2300
CATACATCGT CACCGGMGCG GACGGCATAC TGRKCACCGA GCTCGG 2346

(2) INFORMATION FOR SEQ ID NO: 62

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 841

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#435

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 62

CGTTACCCGC TTTACACCAC CGCCAAGGCC AACCTGACCG CGCTCAGCAC 50
CGGGCTGTCC AGCTGTGCGA TGGCCGACGA CGTGCTGGNC NAGSCCNANS 100
CCAATGNCGG MMTGCTGCAA NCGGNTNCNG GCCANGCGTT CGGACCGGAC 150
GGACGCTGGN CGGTATCAGT CCNGTCGGCT TCAAANCCGA NGGCGTGGGC 200
GAGGACCTCA AGTCCGRRCC CGGTGGTCTC NAAACCCSGG CTNGTCAACT 250
CCGATNCGTC GCCCAACAAN CCCAACNGCC NGCCATCANC GACTCCKCNG 300
GCACCGCCNG AGGGAAGGGY CCGGNTCGGG ATTCAACGGG TTGGCRWCGC 350
GGCGCTGCCG TTCNGRATTG GAYCCGGCAN CGTACCCCGG TGATGGGCAG 400
CTNACGGGGA NGAACAACCY GSCCSSSACG GCCACCTCGG CCTGGTACCA 450
GTTACCGCCC CGCAGCCCGG ACCGGCCNGC TGGTGGTGGT TTCCNGCGGC 500
CGGCGCCATC TGGTCCTACA AGGAGGACGG CGATDTCATC TACGGCCANG 550
TCCCNTGAAA CTGCAGTGGG NCGTCACCGG CCCGGACGGC CGCANTCCAG 600
CCACTGGGGC AGGTATTTCC GANTCGACAN TCGGACCNGC AACNCCNGCG 650
TGGCGCAATC TGCGGTNTNT CCGCTGGCCT GGGCGCCGCC GGNANGCNCG 700
ACGTGGCGCG CATTGTCGCC TATGACCCGA ACCTGAGCCC TGAGCAATGG 750
TTCGCCTTCA CCCCGCCCCG GGTTCCGGTG CTGGAATCTC TGCAGCGGTT 800
GAKCGGGTCA GCGACACCGG TGTTGATGGA CATCGCGACC G 841

(2) INFORMATION FOR SEQ ID NO: 63

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 471

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#1-2/23/9

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 63

GCCAGCCGTG ATCGGCTGAC CGGCAGTGAT CACCAACCTC AACGTGGTGC 50

TGGGCCTCGC TGGCGCTCAC ACGATCGGTT GGACCAGCCG GTGACGTCGC 100

TATCAGCGTT GATTCACCGG CTCGCGCAAC GCAAGACCGA CATCTCCAAC 150

GCCGTGGCCT ACACCAACGC GCCGCCGGCT CGGTCGCCGA TCTCTGTCGC 200

AGGCTCGCGC CGTTGGCGAA GGTGGTTCGC GAGACCGATC GGGTGGCCGG 250

CATCGCGGCC GCCGACCACG ACTACCTCGA CAATCTGCTC AACACGCTGC 300

CGGACAAATA CCAGGCGCTG GTCCGCCAGG GTATGTACGG CGACTTCTTC 350

GCCTTCTACC TGTGCGACGT CGTGCTCAAG GTCAACGGCA AGGGCGGCCA 400

GCCGGTGTAC ATCAAGCTGG CCGGTCAGGA CATGCGGCGG TGCGCGCCGA 450

AATGAAATCC TTCGCCGAAC G 471

(2) INFORMATION FOR SEQ ID NO: 64

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 485

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#1-229/264

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 64

KGTCTCGCGN CCTTNACATC CGGTCGCCNN RCGGTNATCT GCCTGTGGAT 50

GCCGTCCGGA NGTATNANCN AATGGCCANG AGTNCGTGAC NGCAGNTATG 100

GNCKCGGNTA TAGTTCCGTT TTGCCCNGGA CTNGGNGCGT GAGGTGGAAC 150

TAATGGCGGT GTCGGGTGAT ATTTCCGACG GCAAGNCGAC CATATAGGTG 200

GNATNCGACG GCAATAAACA CACGCTCTGG CCACGTTTCT TGGCGGGGAA 250
AGGGGTGATG CTATCGGAGC CAATGGTATC GCGACAACAC TTGCAGATGC 300
CGCCAAGGCC GATCACGCTA ATGACGGATT CGGGGCCACA AACGTTCCCC 350
GTTCTGGCGG TTTTCTCTGA CTACACCTCA GATCAAGGTG TGATTTTGAT 400
GGATCGCGCC AGTTATCGGG CCCATTGGCA GGATGATGAC GTGACGACCA 450
TGTTTCTTTT TTTGGCNATN CGGGTGCGAA TAGCG 485

(2) INFORMATION FOR SEQ ID NO: 65

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 469

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#1-264A

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 65

GGCGAGGTCA GTGAAGCCGA GGAAGCGGAA AGGAGCGCCC AATACGGAAC 50
CGCCTCTCCC CGCGCGTTGG CCGATTCATT AAATGCAGCT GGCACGACAG 100
GTTTCCCGAC TGGAAMGCGG GCAGTGAGCG CAASGCAATT AATGTGAGTT 150
AGCTCACTCA TTAGGCACCC CAGGCTTTAC ACTTTATGCT TCCGGCTCGT 200
ATGTTGTGTG GAATTGTGAG CGGATAACAA TTTCACACAG GAAACAGCTA 250
TGACATGATT ACGAATTTAA TACGACTCAC TATAGGGAAT TCGAGCTCGG 300
TACCCGGGGA TCCTCTAGAG TCGCTTCGGT TGGCGGCGAC CAGCAGTGGA 350
TCCACGGTGG CCGCCCGCGC GGCDTCATAC ACCGCCGCGG CCTCCTTGGC 400
CTGTGCGGCC SGCTTAGCGC GCGTGTTGCT GCCGTGCTTA GCCANCTGGC 450
ATAGGGGGCT GCCGCGCGC 469

(2) INFORMATION FOR SEQ ID NO: 66

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 290

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#1-264C

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 66

CNGGTTCGAC TGATCTAGCT GGGGCCAGAC CGGCACGAGG CGACAGTTAC 50
CAGTACCTGA CAGACAGGCC GATCGAGCCA AACCGTAGTG AGGACGCAGG 100
AGGAACAGGC AGATGCATCT AATGATACCC GCGGAGTATA TCTCCAACGT 150
GATATATGAA GGTCCGCGTG CTGACTCATT GTATGCCGCC GACCAGCGAT 200
TGCGACAATT AGCTGACTCA GTTAGAACGA CTGCCGAGTC GCTCAACACC 250
ACGCTCGACG AGCTGCACGA GAACTGGAAA GGTAGTTTCA 290

(2) INFORMATION FOR SEQ ID NO: 67

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1306

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#2-92

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 67

GTGATACAGG AGGCGCCAAC AGTGACACCT CGCGGGCCAG GTCGTTTGCA 50
ACGCTTGTCG CAGTGCAGGC CTCAGCGCGG CTCCGGAGGG CCTGCCCGTG 100
GTCTTCGACA GCTGGCGCTC GCAGCAATGC TGGGGGCATT GGCCGTCACC 150
GTCAGTGGAT GCAGCTGGTC GGAAGCCCTG GGCATCGGTT GGCCGGAGGG 200
CATTACCCCG GAGGCACACC TCAATCGAGA ACTGTGGATC GGGGCGGTGA 250
TCGCCTCCCT GGCGGTTGGG GTAATCGTGT GGGGTCTCAT CTTCTGGTCC 300
GCGGTATTTC ACCGGAAGAA GAACACCGAC ACTGAGTTGC CCCGCCAGTT 350
CGGCTACAAC ATGCCGCTAG AGCTGGTTCT CACCGTCATA CCGTTCCTCA 400
TCATCTCGGT GCTGTTTTAT TTCACCGTCG TGGTGCAGGA GAAGATGCTG 450
CAGATAGCCA AGGATCCCGA GGTCGTGATT GATATCACGT CTTTCCAGTG 500
GAATTGGAAG TTTGGCTATC AAAGGGTGAA CTTCAAAGAC GGCACACTGA 550
CCTATGATGG TGCCGATCCG GAGCGCAAGC GCGCCATGGT TTCCAAGCCA 600
GAGGGCAAGG ACAAGTACGG CGAAGAGCTG GTCGGGCCGG TGCGCGGGCT 650
CAACACCGAG GACCGGACCT ACCTGAATTT CGACAAGGTC GAGACGTTGG 700
GCACCAGCAC CGAAATTCCG GTGCTGGTGC TGCCGTCCGG CAAGCGTATC 750
GAATTCCAAA TGGCCTCAGC CGATGTGATA CACGCATTCT GGGTGCCGGA 800
GTTCTTGTTC AAGCGTGACG TGATGCCTAA CCCGGTGGCA AACAACTCGG 850
TCAACGTCTT CCAGATCGAA GAAATCACCA AGACCGGAGC ATTCGTGGGC 900

CACTGCGCCG AGATGTGTGG CACGTATCAC TCGATGATGA ACTTCGAGGT 950
CCGCGTCGTG ACCCCCAACG ATTTCAAGGC CTACCTGCAG CAACGCATCG 1000
ACGGGAAKAC AAACGCCGAG GCCCTGCGGG CGATCAACCA GCCGCCCCTT 1050
GCGGTGACCA CCCACCCGTT TGATACTCGC CGCGGTGAAT TGGCCCCGCA 1100
GCCCGTAGGT TAGGACGCTC ATGCATATCG AAGCCCGACT GTTTGAGTTT 1150
GTCGCCGCGT TCTTCGTGGT GACGGCGGTG CTGTACGGCG TGTTGACCTC 1200
GATGTTCGCC ACCGGTGGTG TCGAGTGGGC TGGCACCACT GCGCTGGCGC 1250
TTACCGGCGG CATGGCGTTG ATCGTCGCCA CCTTCTTCCG GTTTGTGGCC 1300
GCGGAT 1306

(2) INFORMATION FOR SEQ ID NO: 68

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 759

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-823

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 68

GGTGCCTGCC ATCGGTTCGC TGNGCCACNG CTGNCNNATC TTTGGTSTGT 50
TAGAGGTNWW CCGCGCGGAT RGCNCANTCC TGTTGGNGGG GGTTRTCGCC 100
ACGATTGCCG CCCGCGCTGA ACCCGACGAC GCCGATGCCC TGCCCACCAC 150
GGATCGGCTG NNMMCANCCG AGCGAACCGT GCAGNATGCN TNTKGTTGAC 200
GAGCCTGCTG GCGCCTTCGC NGGCNCTCGG CGACCATCGG TGCCATCGGA 250
ACCGCCGTNC GCAACCCACG GCATCCACAN GSTCCANGCA TGGCGGTATC 300
GCGNTTGGCC GNCGTCACCG GTGCGCTGCT GCTGCTAYGA GCACGTTCAG 350
CAGACACCAG AAGGTCACTG NTGTTTGCCA TCTGTNGGAA TCACCACCGT 400
TGCAACGGMA NTTGTACCGT CGCCGCGGAT CGGGCTCTGG AACACGGGCC 450
GTGGATTGSC GCGCTGACCG CCATGCTGGT CCNGCCGTGG CAANTGKKTT 500
TGGGCTTCGT NGCTCNCCGC GTTGTCGCTC TCGCCCGTCA CGTACCGCAC 550
CATCGAATTG CTGGAGTGTC TGGCGCTGAT CGCAATGGTT CCATTGACCG 600
CTNTGGSTAT NNNNNCGCCT ANCAGSSSCS TTCGCCACCT CGACCTGACA 650
TGGACATGAC CACNGTCCCG TNACCCTGCG CCTGNCTNGG TGGTMTCAGC 700
GNCNNNTCGY SACGCTGTCT GGSWTGGSRM RCGCNCGGTT GCGCCACGCG 750
GTTTCGCCG 759

(2) INFORMATION FOR SEQ ID NO: 69

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1041

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#1-31

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 69

GKTCNCGGTG ATGTCGACNG TCGGCACGRM GNCGAAACCT CANCGGTCGA 50

CAGTGTCTGC CCGAGGCCGC AGCCGACGTG CCCCNGGAGA CCGCGCGCCA 100

ANCACGGTGC CGTACATGTA GCCCGCACGG CGCATCATCG CCGAGCCGGC 150

GTAGATGTTT TCCTGCACGG CGTNCSCGGT GAACCCTCCG GCGCCAGCAC 200

CGSCACCWNT TCCCGCGTCC ACGTCGGCCT GGGTGGTGAC GCCGAGCACC 250

CCACCGAAAT GATCGACATG GCTGTGGGTG TAGATGACCG SCGACCACGG 300

GGCGGTCGGC TCCGCGGTGG GCGCGANTAC AAGTCCAGCG CGGCGGCGGC 350

CACCTCGGTG GACANCCAAN CGGGYNYGAT GACGARWCWG CCCAGTGTCA 400

CCNCWMMACG AAGNCTGATA TTGGAGATAT CGAATCCGCG GACCTGATAG 450

ATGCCCGGCA CCACCTGGTA GAGGCCCTGT TTCGCGGTCA GCTGGGATTG 500

CCGCCACAGG CTGGGATGCA CCGATGTCGG CGCGGCACCG TCGAGNAACG 550

AGTACGCGTC GTTGTCCCAC ACCNACGCGA CCATCGGCAG CCTTGATCAC 600

ACACGGGGAC AGCGCGGCAA TGAATCCGCG ATCGGCGTCG TCGAAATCCG 650

TTGTGTCATN GCAACGGTNA ACGAGTGTTC ACCGTGTGCC GCCTGGNATG 700

ACGGCAGTNG GGAGGTTTGT GTTCCATCGG CACTACATTG CCACTACTAC 750

GGTGCACGCC GGTAGATGCC GTTGGCGAAC CACGCTACCG ACCAGAAAGA 800

GAGAATTTTC CGCCGCACCT AGACCTCGGG CCCTCTAACG CGCATACTGC 850

CGAAGCGGTC CTCAATGCCG ATGGACCGCT ACGACAGGCA AAGGAGCACA 900

GGGTGAAGCG TGGACTGACG GNTCGCGGTA GCCGGAGCCG CCATTCTGGT 950

CGCAGGTCTT TCCGGATGTT CAAGCAACAA GTCGACTACA GGAAGCGGTG 1000

AGACCACGNA CCGCGNGCAG GCACGACNGC AAGCCCCGGC G 1041

(2) INFORMATION FOR SEQ ID NO: 70

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 799

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#1-3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 70

AGATCNAYAC YANCANCANT GCNGTCATCG AGNTGCTGCA GGNCANGGTG 50
GTCCGTTGGC GAACGTGCTN KGCCNAYACC GGTGCCTTCT CGGCGCNCTN 100
GGYGCAYNGC GACCAGCTGA TCGGCGNAKG TAATCACCAA CCTCAANNKC 150
GGTGCTNGCK ACCKTCGAYK GCAAAGAGYG YGCAATTTGT CGGCCAGTGT 200
CGACCAGCTG CAGCAGCTGG TCAGCGGCCT GGCCAAGAAC CGGGATNCCG 250
ANTSGNGGGC GCCATTTCGC CGCTGGNGTC GACGACGACG GATCTTWCGG 300
AACTGTTGCG GAATTSGCGC CGGCCGCTGC AAGGCAKCCT GGAAAACGCC 350
CGGCCGCTGG CTACCGAGCT GGACAACCGA AAGGCCNANG GTCAASAACG 400
RRATCGAGCA NGCTCGGCGA GGACNATNCC TGCGCCTGTC CGCGCTGGGC 450
AGTTACGGAG CANTTCGTTC AACATCTAST TSTGCTCGGT GACGATSAAG 500
ATCAACGGAC CGGCCGGCAG CGACANTCCN TGCTGCCGAT CGGCGGCCAG 550
CCGGANTCCC AGCAAGGGGA GGTGCGCCTT TGCNTAAATA GGAAGCCAAG 600
TANGCAAASA CGAASGCSAC CCGTCCGCAC CGGNCATCTT CGGCCTGGTG 650
CNTGGTGATC NTGNCGTCGT CCTGATSGNC ATTCGGCTAC AGCGGGTTGC 700
CTKTCTGGCC ACAKKKCAAA ACCTACGACG CGTATTTCAC CGACGCCGGT 750
GGGATCACCC CCGGTAACTC GGTTTATGTS TCGGGCCTCA AGGTGGGCG 799

(2) INFORMATION FOR SEQ ID NO: 71

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 713

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-827 translation strand

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 71

CTAYCSGCAA NGCTKNGCAG ACGCTCGGCT GCACNGCAGA ANTCGCGGTG 50
CACCCACGAT TGCCAGTAGC GCGGGCCCAC TCGTGCCTAC TACACTTCGT 100
CGTAGCCAAA TCANTCGGCC CCGTAGTATC TCCGGAGATG ACAGATGAAT 150
GTCGTCGACA TTTCNGNCGG TGGCAGTTCG GTATCACCAC CGTSTATCAC 200

TTNCAWYTTC GTNACSYGYT GACCWWCGGC CTGGCNCNCC TKSTKANYRC 250
GGNTCNAYGC AAACTGCTGT GGTCGTCACC GATAANCCCG CCTGGTATCG 300
CCTCACCNAA ATTCTTCGGC AAATTGTTCC TGNATCNAAC NTTTGCCATC 350
GGCGTGGCGA CCGGAATCGT GCAGGNAATK TCAGTTCGGC ATGAACTGGA 400
GCGAGTACTC CCGATTCGTC GGCGATGTCT TCGGCGCCCC GCTGGCCATG 450
GAGNSCTGGC GGCCTTNCTT CTTCGAATCC ACCTTCATCG GGTTGTGGAT 500
CTTCGGCTGG AACAGGCTGC CCCGGCTGGT GCANTCTNGG CCTGCATCTG 550
GNATCGTCGC AATNCGCNGG TNCAACGTGT CCGCGTTCTT CATCATCGCN 600
GGCAAACTCC TTCATGCAGC ATCCGGTCGG CGCGCACTAC AACCCGACCA 650
CCGGGCGTGC CGAGTTGAGC AGCATCGNTC NGTGNCNTGC TGACCAACAA 700
CACCGCACAG GCG 713

(2) INFORMATION FOR SEQ ID NO: 72

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 274

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-834 translation strand

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 72

CCGCAGCACC GAGGCAAGCA TCGCACCCGT CGATTCCCGC CATCCCGGCG 50
ACATGATGGT CATGTCCGAC ACCGACGCCC GCACCTCGCT TCCCGAGTTG 100
ACCGCGCTGC GCGTGGACGC CGCAACGGAT GCGTCGGTTC ATTCGATCCC 150
GGCTCGAAAT TGGCCATGGC GAACGCATCT TGCTGTGATG GTTCGGGCAG 200
TAGATCTCCA CTGCCGCACT GATAAACTCG GGTCATGGTC GTCGTGAGGC 250
GGACAGGGTA GAGGCGCATG ACCG 274

(2) INFORMATION FOR SEQ ID NO: 73

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 252

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-874

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 73

GTGATGCCTT CCAGCATTGG ATTGGTCGTC GGTTCGATGC TGTGGCGACA 50

GATAAACCGC CTGTTCGGGG TGCGTGGCCT CTGCTGGGCA GCGCACTGCT 100

CAACGCCGCT CTGCGCTGCT GTGCATGGTG GCCGAGTCGT GTGGGCAGTG 150

GGTTCACGCC TGGGCGTACT TCACGGCGTT CCTGCTGGCT ACGGTGGCCG 200

CTCAAACGGT GGTCGCCGCA TCGATATCGT GGATCAGCGT CCTCGCGCCC 250

GA 252

(2) INFORMATION FOR SEQ ID NO: 74

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 160

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-1018

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 74

GGCGCCGCCG TCGTGCTGGC CGCCCGGCCC GGTGGGGGTG CCGGCCAGCG 50

TGGTTCCGCC AGTGGCCGCG CCGAACGTAT TGGCCGGCGT CCTCGAGCAC 100

GACAACGACG GGTCGGGGGC GGCGGTGCTG GCCGCGCTGG CCAAGCTGCC 150

ACCCGGTGGT 160

(2) INFORMATION FOR SEQ ID NO: 75

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 393

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: **HinPI#1-27

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 75

ATCAGCCGCG GGTCGACGCC GCCGATGACC TCGACGTCGT CGTCGTCGCT 50
GCCGGTACTC AATCCAATCA CCATCCTCTT ACGCACCTTC TAGGAGTGTG 100
TTGCTGCGGC AGTGCCGGCC ATTCGTAGAT TCGGGCCTCG CCGTTGTCGT 150
AGATCTTCGC CCACGACCTC GATGTCTCTA ACGACACTAG TCCGTCCGGC 200
ACGCAAACCC CGCACCGTCG GAGTGCTGGT CAGGTATAGA CGGTACAGGA 250
GGACTTGGTA GGCCTCGAGT ACCGAGGTAC GTCTCCCGTT GCGGCATAGG 300
CCAGAAGATG AACCGGTGTA GACCGGGCCT GTTGCGAGGG TCGTAGTCGT 350
AGGTCCCAGA GGTGTCGGAC GCCCAGGTTA ATACACAGCG TGC 393

(2) INFORMATION FOR SEQ ID NO: 76

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 248

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: #2-147

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 76

GCAGACCTCT GGCCGCTGGT GGTGCTGGGT ACCTGCGCTG GCGACACCGG 50
ACCGCAGACC GTCAATCGGG ACTCCCGGGA ACGTGGTGCC ATCTTGCCAC 100
GGGGATGGCC GACGCGGCTC GTCATTCTCC CCGAGCGCAC CGGCCGCCGC 150
TGTTGACCGG GCCGCGGCGA CTGATGGTGC CCGCACACGC GGGCGGGTTC 200
AAGGAGCAAT ACGCCAAGTC CAGCGCCGCT CTCGCACGGC GCGGTGTT 248

I claim:

1. An isolated *Mycobacterium tuberculosis* nucleic acid sequence including a sequence selected from the group consisting of Seq. I.D. Nos. 1 - 76.

2. A purified immunostimulatory peptide encoded by a sequence according to claim 1.

3. An antibody that specifically binds to a peptide according to claim 2.

4. A vaccine preparation comprising at least one immunostimulatory peptide according to claim 2 and a pharmaceutically acceptable excipient.

5. A purified immunostimulatory peptide encoded by a nucleotide sequence selected from the group consisting of Seq. I.D. Nos. 1 - 76.

6. A vaccine preparation comprising at least one peptide according to claim 5 and a pharmaceutically acceptable excipient.

7. A purified immunostimulatory *Mycobacterium tuberculosis* peptide, the peptide including at least 5 contiguous amino acids encoded by a nucleic acid sequence selected from the group consisting of Seq. I.D. Nos. 1 - 76.

8. A vaccine preparation comprising at least one peptide according to claim 7 and a pharmaceutically acceptable excipient.

9. A peptide according to claim 7 wherein the peptide includes at least 10 contiguous amino acids encoded by a nucleic acid sequence selected from the group consisting of Seq. I.D. Nos. 1 - 76.

10. A vaccine preparation comprising at least one peptide according to claim 9 and a pharmaceutically acceptable excipient.

11. A method of making a vaccine comprising:

providing at least one purified peptide encoded by a nucleotide sequence selected from the group consisting of Seq. ID. Nos 1 - 76;

combining the peptide with a pharmaceutically acceptable excipient.

12. An isolated nucleic acid molecule having a nucleotide sequence selected from the group consisting of:

(a) Seq. ID Nos. 1 - 76;

(b) nucleotide sequences complementary to a sequence defined in (a); and

(c) nucleic acid molecules of at least 15 nucleotides in length which hybridize under conditions of at least 75% stringency to a sequence defined in (a) or (b).

13. A recombinant DNA vector including a nucleic acid molecule according to claim 12.

14. A transformed cell containing a vector according to claim 13.

15. A nucleic acid probe comprising a nucleic acid molecule according to claim 12 and a diagnostic label.

16. A method of isolating a *Mycobacterium tuberculosis* gene which gene encodes an immunostimulatory peptide, the method comprising the steps of:

providing nucleic acids of *Mycobacterium tuberculosis*;

contacting said nucleic acids with a probe or primer, the probe or primer comprising at least 15 contiguous nucleotides of a polynucleotide having a nucleotide sequence selected from the group consisting of Seq. ID Nos. 1 - 76 and sequences complementary thereto; and

isolating the *Mycobacterium tuberculosis* gene.

17. An isolated *Mycobacterium tuberculosis* gene produced by the method of claim 16.

18. An isolated *Mycobacterium tuberculosis* nucleic acid molecule, said molecule encoding an immunostimulatory peptide and hybridizing under conditions of at least 75% stringency to a nucleic acid probe comprising at least 20 contiguous bases of a sequence selected from Seq. ID Nos. 1 - 76.

19. A purified immunostimulatory peptide encoded by the nucleic acid molecule of claim 18.

20. An immunostimulatory preparation comprising:
a purified peptide according to claim 19; and
a pharmaceutically acceptable excipient.

21. An improved tuberculin skin test, the improvement comprising the use of one or more immunostimulatory peptides according to claim 19.

22. A vaccine preparation comprising an immunostimulatory membrane peptide isolated from *Mycobacterium tuberculosis* and a suitable excipient.

23. A method of detecting the presence of *Mycobacterium tuberculosis* DNA in a sample comprising contacting the sample with a nucleic acid probe according to claim 15 and detecting hybridization products that include the nucleic acid probe.

24. A method of detecting the presence of *Mycobacterium tuberculosis* DNA in a sample comprising:

selecting two or more nucleic acid primer molecules from the nucleic acid molecules defined in claim 12, said molecules suitable for amplification of a *Mycobacterium tuberculosis* target sequence;

incubating the sample under conditions suitable to amplify the target sequence; and

detecting an amplified product.

25. A method of detecting the presence of a *Mycobacterium tuberculosis* peptide in a sample comprising contacting the sample with an antibody according to claim 3 and detecting the presence of an antibody-peptide complex.

26. A method of detecting the presence of an anti-*Mycobacterium tuberculosis* antibody in a sample comprising contacting the sample with a peptide according to claim 2 and detecting the presence of an antibody-peptide complex.

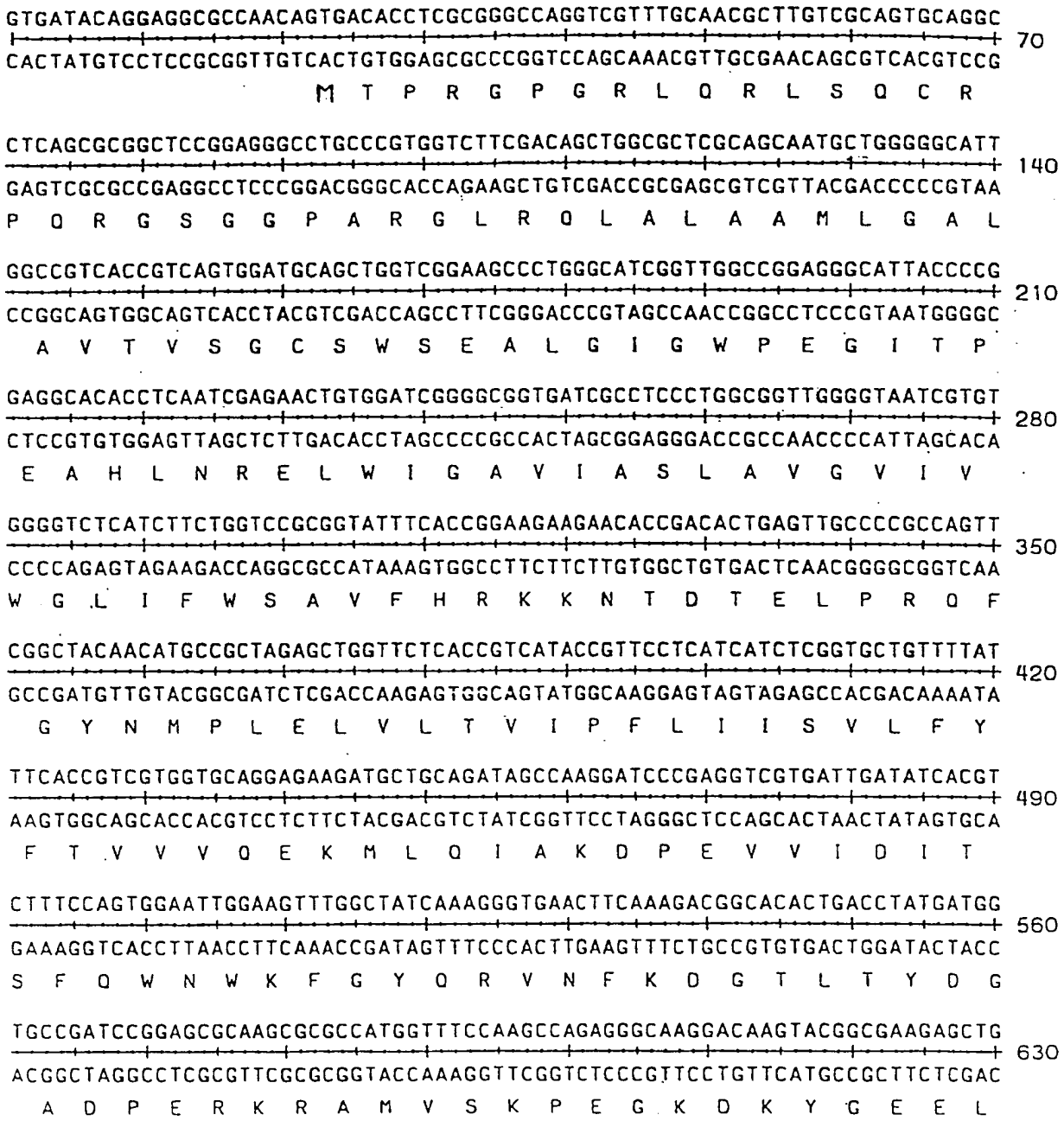

FIG. 1

(Page 1 of 2)

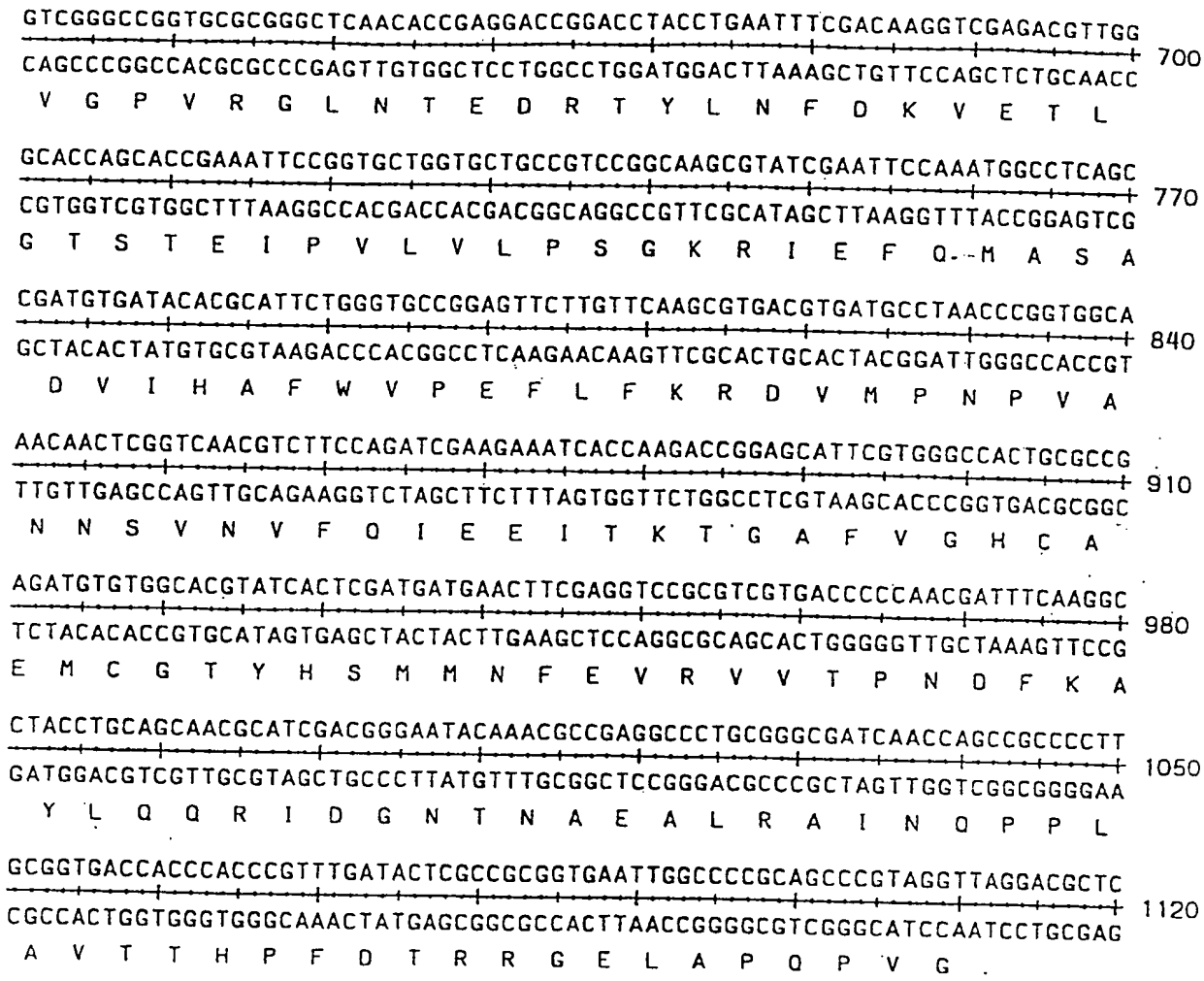

FIG. 1

(Page 2 of 2)

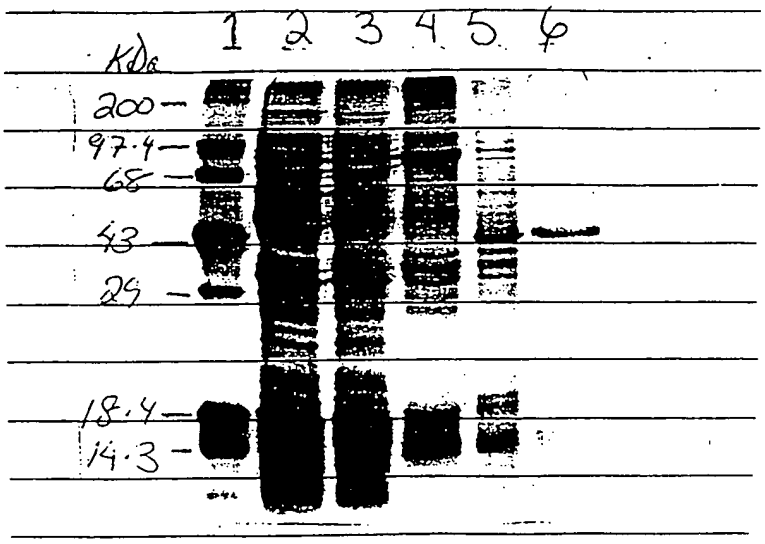

FIG. 2

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/10375

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.
US CL :435/6, 7.1, 240.2, 320.1; 514/2; 530/300, 387.1; 536/23.7, 24.32
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 7.1, 240.2, 320.1; 514/2; 530/300, 387.1; 536/23.7, 24.32

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

STN: Medline, Biosis, CAPlus, WPIDS, JAPIO, PATOSEP, PATOSWO; APS
search terms: mycobacterium tuberculosis, peptide, polypeptide, protein, epitope, antigen, immunostimulat?, membrane, surface

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	TOMMASSEN et al. Use of the enterobacterial outer membrane protein PhoE in the development of new vaccines and DNA probes. Intl. J. Microbiol. Virol. Parasitol. infect. Dis. 1993, VOL. 278, pages 396-406.	1-26
Y	JANSSEN et al. Immunogenicity of a mycobacterial T-cell epitope expressed in outer membrane protein PhoE of Escherichia coli. Vaccine. 1994, Vol.12, pages 406-409.	1-26
Y	Lim et al. Identification of Mycobacterium tuberculosis DNA sequences encoding exported proteins by using phoA gene fusions. J. Bacteriol. January 1995, Vol.177, pages 59-65.	1-26

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
09 OCTOBER 1996	28 OCT 1996
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer ROBERT A. SCHWARTZMAN Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/10375

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

A61K 14/35, 38/16; C07H 16/12, 21/04; C12N 5/10, 15/63; C12Q 1/68; G01N 33/569

CORRECTED VERSION*

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
A61K 14/35, 38/16, C07H 16/12, 21/04, C12N 5/10, 15/63, C12Q 1/68, G01N 33/569

A1

(11) International Publication Number: WO 97/00067

(43) International Publication Date: 3 January 1997 (03.01.97)

(21) International Application Number: PCT/US96/10375

(22) International Filing Date: 14 June 1996 (14.06.96)

(30) Priority Data:
60/000,254 15 June 1995 (15.06.95) US

(71) Applicant *(for all designated States except US)*: UNIVERSITY OF VICTORIA [CA/CA]; 3800 Finnerty Road, Victoria, British Columbia V8W 2Y2 (CA).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: NANO, Francis, E. [US/CA]; #4-245 Ontario Street, Victoria, British Columbia V8V 1N1 (CA).

(74) Agent: POLLEY, Richard, J.; Klarquist Sparkman Campbell Leigh & Whinston, L.L.P., One World Trade Center, Suite 1600, 121 S.W. Salmon Street, Portland, OR 97204-2988 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: MYCOBACTERIUM TUBERCULOSIS DNA SEQUENCES ENCODING IMMUNOSTIMULATORY PEPTIDES

(57) Abstract

Nucleotide sequences isolated from *Mycobacterium tuberculosis* are disclosed. These sequences are shown to encode immunostimulatory peptides. The invention encompasses, among other things, vaccine preparations formulated using these peptides.

* (Referred to in PCT Gazette No. 15/1997, Section II)

Historically it has been thought that one needs replicating *Mycobacteria* in order to effect a protective immunization. An hypothesis explaining the molecular basis for the effectiveness of replicating mycobacteria in inducing protective immunity has been proposed by Orme and co-workers (3). These scientists suggest that antigens are pinocytosed from the mycobacterial-laden phagosome and used in antigen presentation. This hypothesis also explains the basis for secreted proteins effecting a protective immune response.

Antigens that stimulate T cells from *M. tuberculosis* infected mice or from PPD-positive humans are found in both the whole mycobacterial cells and also in the culture supernatants (3, 4, 5-7, 34). Recently Pal and Horwitz (8) were able to induce partial protection in guinea pigs by vaccinating with *M. tuberculosis* supernatant fluids. Similar results were found by Andersen using a murine model of tuberculosis (9). Other studies include reference nos. 34, 12. Although these works are far from definitive they do strengthen the notion that protective epitopes can be found among secreted proteins and that a non-living vaccine can protect against tuberculosis.

For the purposes of vaccine development one needs to find epitopes that confer protection but do not contribute to pathology. An ideal vaccine would contain a cocktail of T-cell epitopes that preferentially stimulate Th1 cells and are bound by different MHC haplotypes. Although such vaccines have never been made there is at least one example of a synthetic T-cell epitope inducing protection against an intracellular pathogen (10). It is an object of this invention to provide *M. tuberculosis* DNA sequences that encode bacterial peptides having an immunostimulatory activity. Such immunostimulatory peptides will be useful in the treatment, diagnosis and prevention of tuberculosis.

II. SUMMARY OF THE INVENTION

The present invention provides DNA sequences isolated from *Mycobacterium tuberculosis*. Peptides encoded by these DNA sequences are shown to stimulate the production of the macrophage-stimulating cytokine, gamma interferon ("INF-γ"), in mice. Critically, the production of INF-γ by CD4 cells in mice has been shown to correlate with maximum expression of protective immunity against tuberculosis (11). Furthermore, in human patients with active "minimal" or "contained" tuberculosis, it appears that the containment of the disease may be attributable, at least in part, to the production of CD4 Th-1-like lymphocytes that release INF-γ (12).

Hence, the DNA sequences provided by this invention encode peptides that are capable of stimulating T-cells to produce INF-γ. That is, these peptides act as epitopes for CD4 T-cells in the immune system. Studies have demonstrated that peptides isolated from an infectious agent and which are shown to be T-cell epitopes can protect against the disease caused by that agent when administered as a vaccine (13, 10). For example, T-cell epitopes from the parasite *Leishmania major* have been shown to be effective when administered as a vaccine (10, 13-14). Therefore, the immunostimulatory peptides (T-cell epitopes) encoded by the disclosed DNA sequences may be used, in purified form, as a vaccine against tuberculosis.

As noted, the nucleotide sequences of the present invention encode immunostimulatory peptides. In a number of instances, these nucleotide sequences are only a part of a larger open reading frame (ORF) of an *M. tuberculosis* operon. The present invention enables the cloning of the complete ORF using standard molecular biology techniques, based on the nucleotide sequences provided herein. Thus, the present invention encompasses both the nucleotide sequences disclosed herein and the complete *M. tuberculosis* ORFs to which they correspond. However, it is noted that since each of the nucleotide sequences disclosed herein encodes an immunostimulatory peptide, the use of larger peptides encoded by the complete ORFs is not necessary for the practice of the invention. Indeed, it is anticipated that, in some instances, proteins encoded by the corresponding ORFs may be less immunostimulatory than the peptides encoded by the nucleotide sequences provided herein.

One aspect of the present invention is an immunostimulatory preparation comprising at least one peptide encoded by the DNA sequences presented herein. Such a preparation may include the purified peptide or peptides and one or more pharmaceutically acceptable adjuvants, diluents and/or excipients. Another aspect of the

MYCOBACTERIUM TUBERCULOSIS DNA SEQUENCES ENCODING IMMUNOSTIMULATORY PEPTIDES

CROSS REFERENCE TO RELATED CASES

This application claims the benefit of U.S. Provisional Application No. 60/000,254, filed June 15, 1995, which is incorporated herein by reference.

I. BACKGROUND

A. THE RISE OF TUBERCULOSIS

Over the past few years the editors of the Morbidity and Mortality Weekly Report have chronicled the unexpected rise in tuberculosis cases. It has been estimated that worldwide there are one billion people infected with *M. tuberculosis*, with 7.5 million active cases of tuberculosis. Even in the United States, tuberculosis continues to be a major problem especially among the homeless, Native Americans, African-Americans, immigrants, and the elderly. HIV-infected individuals represent the newest group to be affected by tuberculosis. Of the 88 million new cases of tuberculosis expected in this decade approximately 10% will be attributable to HIV infection.

The emergence of multi-drug resistant strains of *M. tuberculosis* has complicated matters further and even raises the possibility of a new tuberculosis epidemic. In the U.S. about 14% of *M. tuberculosis* isolates are resistant to at least one drug, and approximately 3% are resistant to at least two drugs. *M. tuberculosis* strains have even been isolated that are resistant to all seven drugs in the repertoire of drugs commonly used to combat tuberculosis. Resistant strains make treatment of tuberculosis extremely difficult: for example, infection with *M. tuberculosis* strains resistant to isoniazid and rifampin leads to mortality rates of approximately 90% among HIV-infected individuals. The mean time to death after diagnosis in this population is 4-16 weeks. One study reported that of nine immunocompetent health care workers and prison guards infected with drug resistant *M. tuberculosis*, five died. The expected mortality rate for infection with drug sensitive *M. tuberculosis* is 0%.

The unrelenting persistence of mycobacterial disease worldwide, the emergence of a new, highly susceptible population, and the recent appearance of drug resistant strains point to the need for new and better prophylactic and therapeutic treatments of mycobacterial diseases.

B. TUBERCULOSIS AND THE IMMUNE SYSTEM

Infection with *M. tuberculosis* can take on many manifestations. The growth in the body of *M. tuberculosis* and the pathology that it induces is largely dependent on the type and vigor of the immune response. From mouse genetic studies it is known that innate properties of the macrophage play a large role in containing disease (1). Initial control of *M. tuberculosis* may also be influenced by reactive $\gamma\delta$ T cells. However, the major immune response responsible for containment of *M. tuberculosis* is via helper T cells (Th1) and to a lesser extent cytotoxic T cells (2). Evidence suggests that there is very little role for the humoral response. The ratio of responding Th1 to Th2 cells has been proposed to be involved in the phenomenon of suppression.

Th1 cells are thought to convey protection by responding to *M. tuberculosis* T cell epitopes and secreting cytokines, particularly interferon-γ, which stimulate macrophages to kill *M. tuberculosis*. While such an immune response normally clears infections by many facultative intracellular pathogens, such as *Salmonella*, *Listeria* or *Francisella*, it is only able to contain the growth of other pathogens such as *M. tuberculosis* and *Toxoplasma*. Hence, it is likely that *M. tuberculosis* has the ability to suppress a clearing immune response, and mycobacterial components such as lipoarabinomannan are thought to be potential agents of this suppression. Dormant *M. tuberculosis* can remain in the body for long periods of time and can emerge to cause disease when the immune system wanes due to age or other effects such as infection with HIV-1.

"Probes" and "primers". Nucleic acid probes and primers may readily be prepared based on the nucleic acid sequences provided by this invention. A "probe" comprises an isolated nucleic acid attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. Methods for labeling and guidance in the choice of labels appropriate for various purposes are discussed, e.g., in reference nos. 15 and 16.

"Primers" are short nucleic acids, preferably DNA oligonucleotides 15 nucleotides or more in length, which are annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, then extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods known in the art.

As noted, probes and primers are preferably 15 nucleotides or more in length, but, to enhance specificity, probes and primers of 20 or more nucleotides may be preferred.

Methods for preparing and using probes and primers are described, for example, in reference nos. 15, 16 and 17. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, © 1991, Whitehead Institute for Biomedical Research, Cambridge, MA).

"Substantial similarity". A first nucleic acid is "substantially similar" to a second nucleic acid if, when optimally aligned (with appropriate nucleotide insertions or deletions) with the other nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 75%-90% of the nucleotide bases, and preferably greater than 90% of the nucleotide bases. ("Substantial sequence complementarity" requires a similar degree of sequence complementarity.) Sequence similarity can be determined by comparing the nucleotide sequences of two nucleic acids using sequence analysis software such as the Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, Madison, WI).

"Operably linked". A first nucleic acid sequence is "operably" linked with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Recombinant". A "recombinant" nucleic acid is one that has a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques.

"Stringent Conditions" and "Specific". The nucleic acid probes and primers of the present invention hybridize under stringent conditions to a target DNA sequence, e.g., to a full length *Mycobacterium tuberculosis* gene that encodes an immunostimulatory peptide.

The term "stringent conditions" is functionally defined with regard to the hybridization of a nucleic-acid probe to a target nucleic acid (i.e., to a particular nucleic acid sequence of interest) by the hybridization procedure discussed in Sambrook et al. (1989) (reference no. 15) at 9.52-9.55. *See also*, reference no. 15 at 9.47-9.52, 9.56-9.58; reference no. 18 and reference no. 19.

Nucleic-acid hybridization is affected by such conditions as salt concentration, temperature, or organic solvents, in addition to the base composition, length of the complementary strands, and the number of nucleotide-base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art.

In preferred embodiments of the present invention, stringent conditions are those under which DNA molecules with more than 25% sequence variation (also termed "mismatch") will not hybridize. Such conditions

invention is a vaccine comprising one or more peptides encoded by nucleotide sequences provided herein. This vaccine may also include one or more pharmaceutically acceptable excipients, adjuvants and/or diluents.

Another aspect of the present invention is an antibody specific for an immunostimulatory peptide encoded by a nucleotide sequence of the present invention. Such antibodies may be used to detect the present of *M. tuberculosis* antigens in medical specimens, such as blood or sputum. Thus, these antigens may be used to diagnose tuberculosis infections.

The present invention also encompasses the diagnostic use of purified peptides encoded by the nucleotide sequences of the present invention. Thus, the peptides may be used in a diagnostic assay to detect the presence of antibodies in a medical specimen, which antibodies bind to the *M. tuberculosis* peptide and indicate that the subject from which the specimen was removed was previously exposed to *M. tuberculosis*.

The present invention also provides an improved method of performing the tuberculin skin test to diagnose exposure of an individual to *M. tuberculosis*. In this improved skin test, purified immunostimulatory peptides encoded by the nucleotide sequences of this invention are employed. Preferably, this skin test is performed with one set of the immunostimulatory peptides, while another set of the immunostimulatory peptides is used to formulate vaccine preparations. In this way, the tuberculin skin test will be useful in distinguishing between subjects infected with tuberculosis and subjects who have simply been vaccinated. In this manner, the present invention may overcome a serious limitation inherent in the present BCG vaccine/tuberculin skin test combination.

Other aspects of the present invention include the use of probes and primers derived from the nucleotide sequences disclosed herein to detect the presence of *M. tuberculosis* nucleic acids in medical specimens.

A further aspect of the present invention is the discovery that a significant proportion of the immunostimulatory peptides are homologous to proteins known to be located in bacterial cell surface membranes. This discovery suggests that membrane-bound peptides, particularly those from *M. tuberculosis*, may be a new source of antigens for use in vaccine preparations.

III. BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows the deduced amino acid sequence of the full length MTB2-92 protein.

Fig. 2 shows an SDS polyacrylamide gel (12%) representing the different stages of the purification of MTB2-92 Lane 1:- Molecular weight markers (high range, GIBCO-BRL, Grand Island, NY, U.S.A.); Lane 2:- the IPTG induced crude bacterial lysate of *E. coli* JM109 containing pMAL-MTB2-92; Lane 3:- Uninduced crude bacterial lysate of *E. coli* JM109 containing pMAL-MTB2-92; Lane 4:- Eluate from the amylose-resin column containing the MBP-MTB2-92 fusion protein; Lane 5:- Eluate shown in previous lane after cutting with protease Factor Xa; Lane 6:- Eluate from the Ni-NTA column, containing MTB2-92.

IV. DESCRIPTION OF THE INVENTION

A. DEFINITIONS

Particular terms and phrases used herein have the meanings set forth below.

"Isolated". An "isolated" nucleic acid has been substantially separated or purified away from other nucleic acid sequences in the cell of the organism in which the nucleic acid naturally occurs, i.e., other chromosomal and extrachromosomal DNA and RNA. The term "isolated" thus encompasses nucleic acids purified by standard nucleic acid purification methods. The term also embraces nucleic acids prepared by recombinant expression in a host cell as well as chemically synthesized nucleic acids.

The nucleic acids of the present invention comprise at least a minimum length able to hybridize specifically with a target nucleic acid (or a sequence complementary thereto) under stringent conditions as defined below. The length of a nucleic acid of the present invention is preferably 15 nucleotides or greater in length, although a shorter nucleic acid may be employed as a probe or primer if it is shown to specifically hybridize under stringent conditions with a target nucleic acid by methods well known in the art.

plating clones on agar plates containing the indicator 5-bromo-4-chloro-3-indolyl-phosphate. Alkaline phosphatase converts this indicator to a blue colored product. Hence, those clones containing secreted alkaline phosphatase fusion proteins will produce the blue color.

The three vectors in this series (pJDT1, 2 and 3) have the *Bst*BI restriction sites located in different reading frames with respect to the *pho*A gene. This increases the likelihood of cloning any particular gene in the correct orientation and reading frame for expression by a factor of 3. Reference no. 31 describes pJDT vectors in detail.

3. SELECTION OF SECRETED FUSION PROTEINS

The recombinant clones described above were transformed into *E. coli* and plated on agar plates containing the indicator 5-bromo-4-chloro-3-indolyl-phosphate. Production of blue pigmentation, produced as a result of the action of alkaline phosphatase on the indicator, indicated the presence of secreted cytoplasmic membrane periplasmic, cell wall associated or outer membrane fusion proteins (because the bacterial alkaline phosphatase gene in the vector lacks a signal sequence and could not otherwise escape the bacterial cell). A similar technique has been used to identify *M. tuberculosis* genes encoding exported proteins by Lim et al. (32).

Those clones producing blue pigmentation were picked and grown in liquid culture to facilitate the purification of the alkaline phosphatase fusion proteins. These recombinant clones were designated according to the restriction enzyme used to digest the *Mycobacterium tuberculosis* DNA (thus, clones designated A#2-1, A#2-2 etc were produced using *Mycobacterium tuberculosis* DNA digested with *Aci*I).

4. PURIFICATION OF SECRETED FUSION PROTEINS

PhoA fusion proteins were extracted from the selected *E. coli* clones by cell lysis and purified by SDS polyacrylamide gel electrophoresis. Essentially, individual *E. coli* clones are grown overnight at 30°C with shaking in 2 ml LB broth containing ampicillin, kanamycin and IPTG. The cells are precipitated by centrifugation and resuspended in 100 μL Tris -EDTA buffer. 100 μL lysis buffer (1% SDS, 1mMEDTA, 25mM DTT, 10% glycerol and 50 mM tris-HCl, pH 7.5) is added to this mixture and DNA released from the cells is sheared by passing the mixture through a small gauge syringe needle. The sample is then heated for 5 minutes at 100°C and loaded onto an SDS PAGE gel (12 cm x 14 cm x 1.5 mm, made with 4% (w/v) acrylamide in the stacking section and 10% (w/v) acrylamide in the separating section). Several samples from each clone are loaded onto each gel.

The samples are electrophoresed by application of 200 volts to the gel for 4 hours. Subsequently, the proteins are transferred to a nitrocellulose membrane by Western blotting. A strip of nitrocellulose is cut off to be processed with antibody, and the remainder of the nitrocellulose is set aside for eventual elution of the protein. The strip is incubated with blocking buffer and then with anti-alkaline phosphatase primary antibody, followed by incubation with anti-mouse antibody conjugated with horse radish peroxidase. Finally, the strip is developed with the NEN DuPont Renaissance kit to generate a luminescent signal. The migratory position of the PhoA fusion protein, as indicated by the luminescent label, is measured with a ruler, and the corresponding region of the undeveloped nitrocellulose blot is excised.

This region of nitrocellulose, which contains the PhoA fusion protein, is then incubated in 1 ml 20% acetronitrile at 37°C for 3 hours. Subsequently, the mixture is centrifuged to remove the nitrocellulose and the liquid is transferred to a new test tube and lyophilized. The resulting protein pellet is dissolved in 100 μL of endotoxin-free, sterile water and precipitated with acetone at -20°C. After centrifugation the bulk of the acetone is removed and the residual acetone is allowed to evaporate. The protein pellet is re-dissolved in 100 μL of sterile phosphate buffered saline. This procedure can be scaled up by modification to include IPTG induction 2 hours prior to cell harvesting, washing nitrocellulose membranes with PBS prior to acetonitrile extraction and lyophilization of acetonitrile extracted and acetone precipitated protein samples.

are also referred to as conditions of 75% stringency (since hybridization will occur only between molecules with 75% sequence identity or greater). In more preferred embodiments, stringent conditions are those under which DNA molecules with more than 15% mismatch will not hybridize (conditions of 85% stringency). In most preferred embodiments, stringent conditions are those under which DNA molecules with more that 10% mismatch will not hybridize (i.e. conditions of 90% stringency).

When referring to a probe or primer, the term "specific for (a target sequence)" indicates that the probe or primer hybridizes under stringent conditions substantially only to the target sequence in a given sample comprising the target sequence.

"Purified" - a "purified" peptide is a peptide that has been extracted from the cellular environment and separated from substantially all other cellular peptides. As used herein, the term peptide includes peptides, polypeptides and proteins. In preferred embodiments, a "purified" peptide is a preparation in which the subject peptide comprises 80% or more of the protein content of the preparation. For certain uses, such as vaccine preparations, even greater purity may be necessary.

"Immunostimulatory" - the phrase "immunostimulatory peptide" as used herein refers to a peptide that is capable of stimulating INF-γ production in the assay described in section B 5 below. In preferred embodiments, an immunostimulatory peptide is one capable of inducing greater than twice the background level of this assay determined using T-cells stimulated with no antigens or negative control antigens. Preferably, the immunostimulatory peptides are capable of inducing more than 0·01 ng/ml of INF-γ in this assay system. In more preferred embodiments, an immunostimulatory peptide is one capable of inducing greater than 10 ng/ml of INF-γ in this assay system.

B. MATERIALS AND METHODS

1. STANDARD METHODOLOGIES

The present invention utilizes standard laboratory practices for the cloning, manipulation and sequencing of nucleic acids, purification and analysis of proteins and other molecular biological and biochemical techniques, unless otherwise stipulated. Such techniques are explained in detail in standard laboratory manuals such as Sambrook et al. (15); and Ausubel et al. (16).

Methods for chemical synthesis of nucleic acids are discussed, for example, in reference nos. 20 and 21. Chemical synthesis of nucleic acids can be performed, for example, on commercial automated oligonucleotide synthesizers.

2. ISOLATION OF *MYCOBACTERIUM TUBERCULOSIS* DNA SEQUENCES ENCODING IMMUNOSTIMULATORY PROTEINS

Mycobacterium tuberculosis DNA was obtained by the method of Jacobs et al. (22). Samples of the isolated DNA were partially digested with one of the following restriction enzymes *Hin*PI, *Hpa*II, *Aci*I, *Taq*I, *Bsa*HI, *Nar*I. Digested fragments of 0·2-5kb were purified from agarose gels and then ligated into the *Bst*BI site in front of the truncated *phoA* gene in one or more of the three phagemid vectors pJDT1, pJDT2, and JDT3.

A schematic representation of the phagemid vector pJDT2 is provided in Mdluli et al. (1995) (reference no. 31). The pJDT vectors were specifically designed for cloning and selecting genes encoding cell wall-associated, cytoplasmic membrane associated, periplasmic or secreted proteins (and especially for cloning such genes from GC rich genomes, such as the *Mycobacterium tuberculosis* genome). The vectors have a *Bst*BI cloning site in frame with the bacterial alkaline phosphatase gene (*phoA*) such that cloning of an in-frame sequence into the cloning site will result in the production of a fusion protein. The *phoA* gene encodes a version of the alkaline phosphatase that lacks a signal sequence; hence, only if the DNA cloned into the *Bst*BI site includes a signal sequence or a transmembrane sequence can the fusion protein be secreted to the medium or inserted into cytoplasmic membrane, periplasm or cell wall. Those clones encoding such fusion proteins may be detected by

Cytokines are measured using an Enzyme Linked Immunosorbent Assay (ELISA), the details of which are described in the Cytokine ELISA Protocol in the PharMingen catalogue (PharMingen, San Diego, California). For measuring for the presence of human gamma-interferon, wells of a 96 well microtitre plate are coated with a capture antibody (anti-human gamma-interferon antibody). The sample supernatants are then added to individual wells. Any gamma-interferon present in the sample will bind to the capture antibody. The wells are then washed. A detection antibody (anti-human gamma-interferon antibody), conjugated to biotin, is added to each well, and will bind to any gamma-interferon that is bound to the capture antibody. Any unbound detection antibody is washed away. An avidin peroxidase enzyme is added to each well (avidin binds tightly to the biotin on the detection antibody). Any excess unbound enzyme is washed away. Finally, a chromogenic substrate for the enzyme is added and the intensity of the colour reaction that occurs is quantitated using an ELISA plate reader. The quantity of the gamma-interferon in the sample supernatants is determined by comparison with a standard curve using known quantities of human gamma-interferon.

Measurement of other cytokines, such as Interleukin-2 and Interleukin-4, can be determined using the same protocol, with the appropriate substitution of reagents (monoclonal antibodies and standards).

7. DNA SEQUENCING

The sequencing of the alkaline phosphatase fusion clones was undertaken using the AmpliCycle thermal sequencing kit (Perkin Elmer, Applied Biosystems Division, 850 Lincoln Centre Drive, Foster City, CA 94404, U.S.A.), using a primer designed to read out of the alkaline phosphatase gene into the *Mycobacterium tuberculosis* DNA insert, or primers specific to the cloned sequences.

C. RESULTS

1. IMMUNOSTIMULATORY CAPACITY

More than 300 fusion clones were tested for their ability to stimulate INF-γ production. Of these, 80 clones were initially designated to have some ability to stimulate INF-γ production. Tables 1 and 2 show the data obtained for these 80 clones. Clones placed in Table 1 showed the greatest ability to stimulate INF-γ production (greater than 10 ng/ml of INF-γ) while clones placed in Table 2 stimulated the production of between 2 ng/ml and 10 ng/ml of INF-γ. Background levels of INF-γ production (i.e., levels produced without any added *M. tuberculosis* antigen) were subtracted from the levels produced by the fusions to obtain the figures shown in these tables.

TABLE 1

Immunostimulatory AP-fusion clones

No.	Name	INF	Fus-MW	TBport	coding	Similarity (score)
1	AciI#1-152	>40,000	~65,000	~23,400	~633	*M. avium* acetolactate synthase (98*)
2	AciI#1-247	>40,000	~160,000	~118,400	~3,198	peptide synthetase (153)
3	AciI#1-264	>40,000	~72,500	~30,900	~833	nothing evident
4	AciI#1-435	>40,000	~80,000	~38,400	~1,038	*M. smegmatis* ethambutol resistance gene EmbA (624)

5. DETERMINATION OF IMMUNOSTIMULATORY CAPACITY IN MICE

The purified alkaline phosphatase - *Mycobacterium tuberculosis* fusion peptides encoded by the recombinant clones were then tested for their ability to stimulate INF-γ production in mice. The test used to determine INF-γ stimulation is as essentially that described by Orme et al. (11).

Essentially, the assay method is as follows: The virulent strain *M. tuberculosis* Erdman is grown in Proskauer Beck medium to mid-log phase, then aliquoted and frozen at -70°C for use as an inoculant. Cultures of this bacterium are grown and harvested and mice are inoculated with 1 x 10^5 viable bacteria suspended in 200 µl sterile saline via a lateral tail vein on day one of the test.

Bone marrow-derived macrophages are used in the test to present the bacterial alkaline phosphatase-*Mycobacterium tuberculosis* fusion protein antigens. These macrophages are obtained by harvesting cells from mouse femurs and culturing the cells in Dulbecco's modified Eagle medium as described by Orme et al. (11). Eight to ten days later, up to ten µg of the fusion peptide to be tested is added to the macrophages and the cells are incubated for 24 hours.

The CD4 cells are obtained by harvesting spleen cells from the infected mice and then pooling and enriching for CD4 cells by removal of adherent cells by incubation on plastic Petri dishes, followed by incubation for 60 minutes at 37°C with a mixture of J11d.2, Lyt-2.43, and GL4 monoclonal antibody (mAb) in the presence of rabbit complement to deplete B cells and immature T cells, CD8 cells, and γδ cells, respectively. The macrophages are overlaid with 10^6 of these CD4 cells and the medium is supplemented with 5 U IL-2 to promote continued T cell proliferation and cytokine secretion. After 72 hours, cell supernatants are harvested from sets of triplicate wells and assayed for cytokine content.

Cytokine levels in harvested supernatants are assayed by sandwich ELISA as described by Orme et al. (11).

6. DETERMINATION OF IMMUNOSTIMULATORY CAPACITY IN HUMANS

The purified alkaline phosphatase - *Mycobacterium tuberculosis* fusion peptides encoded by the recombinant clones or by synthetic peptides are tested for their ability to induce INF-γ production by human T cells in the following manner.

Blood from tuberculin positive people (producing a tuberculin positive skin test) is collected in EDTA coated tubes, to prevent clotting. Mononuclear cells are isolated using a modified version of the separation procedure provided with the NycoPrep™ 1.077 solution (Nycomed Pharma AS, Oslo, Norway). Briefly, the blood is diluted in an equal volume of a physiologic solution, such as Hanks Balanced Salt solution (HBSS), and then gently layered over top of the Nycoprep solution in a 2 to 1 ratio in 50 ml tubes. The tubes are centrifuged at 800 x g for 20 minutes and the mononuclear cells are then removed from the interface between the Nycoprep solution and the sample layer. The plasma is removed from the top of the tube and filtered through a 0.2 micron filter and is then added to the tissue culture media. The mononuclear cells are washed twice: the cells are diluted in a physiologic solution, such as HBSS or RPMI 1640, and centrifuged at 400 x g for 10 minutes. The mononuclear cells are then resuspended to the desired concentration in tissue culture media (RPMI 1640 containing 10% autologous serum, Hepes, non-essential amino acids, antibiotics and polymixin B). The mononuclear cells are then cultured in 96 well microtitre plates.

Peptides or PhoA fusion proteins are then added to individual wells in the 96 well plate, and cells are then placed in an incubator (37°C, 5% CO_2). Samples of the supernatants (tissue culture media from the wells containing the cells) are collected at various time points (from 3 to 8 days) after the addition of the peptides or PhoA fusion proteins. The immune responsiveness of T cells to the peptides and PhoA fusion proteins is assessed by measuring the production of cytokines (including gamma-interferon).

TABLE 1

Immunostimulatory AP-fusion clones

No.	Name	INF	Fus-MW	TBport	coding	Similarity (score)
21	AciI#2-1084	>20,000	~73,000	~31,400	~849	Sequences within *M. tuberculosis* clone X68281 (96⁺) and *M. leprae* clone B983 (122⁺)
22	AciI#3-47	>20,000	~55,000	~13,400	~363	nothing evident
23	AciI#3-133	>20,000	~55,000	~13,400	~363	nothing evident
24	AciI#3-166	>20,000	~48,000	~6,400	~174	nothing evident
25	AciI#3-167	>20,000	~65,000	~23,400	~633	*M. leprae* DNA sequence within region B983 (588⁺)
26	AciI#3-206	>20,000	~65,000	~23,400	~633	*M. leprae* DNA sequence within chromosome region MD0092 (91)
27	HinP#1-31	14,638	~46,000	~4,400	~120	*M. tuberculosis* 19 kDa lipo-protein antigen precursor (218)
28	HinP#1-144	13,546	~70,000	~23,900	~645	*M. leprae* DNA sequence within chromosome region B983 (78)
29	HinP#1-3	11,550	~49,000	~7,400	~200	*M. leprae* DNA sequence within chromosome region B983 (100⁺)
30	AciI#1-486	11,416	~45,000	~3,400	~93	nothing known
31	AciI#1-426	11,135	~47,500	~5,900	~160	Dipeptide transport protein (65)
32	AciI#2-916	10,865	~75,000	~33,400	~903	nothing evident

Abbreviations: INF: pg/ml of INF-γ produced using fusion to stimulate immune T-cells. Fus. MW: Relative molecular weight of the fusion protein in Da. TB port.: Estimated amount of fusion attributable to the *M. tuberculosis* protein. Coding: Amount of DNA needed to encode TB portion of fusion proteins (in base pairs). Similarity: Amino acid sequence similarity seen by analysis of DNA via the BLASTX or TBLASTX programs. Scores for alignments are indicated in (). Due to the high G+C nature of M. TB DNA many false positives are evident. Only scores above 100 have good credibility.

TABLE 1

Immunostimulatory AP-fusion clones

No.	Name	INF	Fus-MW	TBport	coding	Similarity (score)
5	HinP#1-27	>20,000	59,000	17,400	471	nothing evident
6	HinP#2-92	>20,000	74,600	33,000	891	1. *M. tuberculosis* ORF MTCY190.11C (1794+) 2. Cytochrome C oxidase subunit II (141)
7	HinP#2-145	>20,000	60,000	13,900	375	nothing evident
8	HinP#2-150	>20,000	55,000	13,400	362	nothing evident
9	HinP#1-200	>20,000	53,500	11,900	321	nothing evident
10	HinP#3-30	>20,000	69,000	27,400	740	*M. leprae* chromosome sequence in B983 region (281+)
11	AciI#2-2	>20,000	70,000	28,400	768	*M. leprae* chromosome sequence within region B1529 (139)
12	AciI#2-23	>20,000	75,000	33,400	903	Region within sequence MD0009 of the *M. leprae* chromosome
13	AciI#2-506	>20,000	60,000	18,400	498	nothing evident
14	AciI#2-511	>20,000	~60,000	~18,400	~498	nothing evident
15	AciI#2-639	>20,000	~60,000	~18,400	~498	nothing evident
16	AciI#2-822	>20,000	~45,000	~3,400	~93	*M. tuberculosis* sequence within region MD0074 (U27357) (551-)
17	AciI#2-823	>20,000	~46,500	~4,900	~132	nothing evident
18	AciI#2-825	>20,000	~150,000	~110,000	~2,970	*M. tuberculosis* sequence MTCY31.03c (431)
19	AciI#2-827	>20,000	~48,000	~6,400	~174	cytochrome d oxidase
20	AciI#2-898	>20,000	~49,000	~7,400	~201	nothing evident

TABLE 2

Immunostimulatory AP-fusion clones (cont'd)

No.	Clone Name	INF	Fus-MW	TBport	coding	Similarity (score)
20	AciI#2-1035	3,454	~46,000	~4,400	~120	nothing evident
21	AciI#2-1089	8,974	~65,000	~23,400	~633	Similar to *M. tuberculosis* sequence X75361 and sequence in *M. bovis* MD0057 and U34849 regions. Immunogenic proteins MPB64 and MPT64 are homologous.
22	AciI#2-1090	7,449	~65,000	~23,400	~633	nothing evident
23	AciI#2-1104	5,148	~68,000	~26,400	~714	Similar to *M. tuberculosis* sequence X80268 and to cds 1 (256) in *M. leprae* sequence region MD0045 (169*); secreted antigenic protein.
24	AciI#3-9	3,160	~67,000	~25,400	~687	nothing evident
25	AciI#3-12	3,891	~75,000	~33,400	~903	Penicillin binding protein; similar to *M. leprae* sequence within genomic clone B1529
26	AciI#3-15	4,019	~65,000	~23,400	~633	nothing evident
27	AciI#3-21	2,301	~69,000	~27,400	~741	nothing evident
28	AciI3-78	2,905	~65,000	~23,400	~633	Similar to sequence within *M. leprae* genomic clone B983
29	AciI#3-134	3,895	~45,000	~3,400	~93	nothing evident
30	AciI#3-204	4,774	~60,000	~13,900	~375	nothing evident
31	AciI#3-214	7,333	~50,000	8,400	~228	nothing evident
32	AciI#3-243	2,857	~65,000	~23,400	~633	nothing evident
33	AciI#3-281	2,943	~65,000	~23,400	~633	Similar to sequence within *M. leprae* genomic clone B983
34	Bsa HI#1-21	8,122	~90,000	~48,400	~1,209	nothing evident
35	HinP#1-12	2,905	~66,000	~24,400	~660	possible tyrosine phosphatase

TABLE 2

Immunostimulatory AP-fusion clones (cont'd)

No.	Clone Name	INF	Fus-MW	TBport	coding	Similarity (score)
1	AciI#1-62	3,126	~43,000	~1,400	~39	*M. tuberculosis* MTCY 190.11C cytochrome C oxidase subunit II (198) *M. leprae* sequence in B1551 region (1087+)
2	AciI#2-14	6,907	~45,000	~3,400	~93	nothing evident
3	AciI#2-26	3,089	~72,000	~30,400	~822	nothing evident
4	AciI#2-35	3,907	~45,000	~3,400	~93	Possibly similar to *M. leprae* sequence in the B983 region (116-)
5	AciI#2-147	5,464				nothing evident
6	AciI#2-508	7,052	~70,000	~28,400	~768	Similar to sequence of the *M. leprae* ORF encoding gp U00018 (125) and similar to sequence in the B2168 c2-209 region of *M. leprae* genome (225+)
7	AciI#2-510	2,445	~69,000	~27,400	~741	nothing evident
8	AciI#2-523	2,479	~50,000	~8,400	~228	Similar to *M. tuberculosis* sequence z70692 from clone Y427 (96)
9	AciI#2-676	3,651	~70,000	~28,400	~768	Similar to AciI#2-639
10	AciI#2-834	5,942	~60,000	~13,900	~375	nothing evident
11	AciI#2-854	5,560	~44,000	~2,400	~66	nothing evident
12	AciI#2-872	2,361	~47,000	~5,400	~147	nothing evident
13	AciI#2-874	2,171	~45,000	~3,400	~93	nothing evident
14	AciI#2-8841	2,729	~85,000	~43,400	~1173	Isocitrate dehydrogenase (247)
15	AciI#2-894	3,396	~70,000	~28,400	~768	nothing evident
16	AciI#2-1014	6,302	~45,000	~3,400	~93	nothing evident
17	AciI#2-1018	4,642	~55,000	~13,400	~363	nothing evident
18	AciI#2-1025	3,582	~45,000	~3,400	~93	nothing evident
19	AciI#2-1034	2,736	~80,000	~38,400	~103	nothing evident

TABLE 2

Immunostimulatory AP-fusion clones (cont'd)

No.	Clone Name	INF	Fus-MW	TBport	coding	Similarity (score)
46	HpaII#1-8	2,048	110,000	68,400	~1,848	nothing evident
47	HpaII#1-10	4,178	55,000	13,400	~633	Similar to immunogenic proteins MPB64/MPT64
48	HpaII#1-13	3,714	43,000	1,400	~39	nothing evident

Abbreviations: INF: pg/ml of INF-γ produced using fusion to stimulate immune T-cells. Fus. MW: Relative molecular weight of the fusion protein. TB port.: Estimated amount of fusion attributable to the *M. tuberculosis* protein. Coding: Amount of DNA needed to encode TB portion of fusion proteins. Similarity: Amino acid sequence similarity seen by analysis of DNA via the BLASTX or TBLASTX' programs. Scores for alignments are indicated in (). Due to the high G+C nature of M. TB DNA many false positives are evident. Only scores above 100 have good credibility.

2. DNA SEQUENCING AND DETERMINATION OF OPEN READING FRAMES

DNA sequence data for the sequences of the *Mycobacterium tuberculosis* DNA present in the clones shown in Tables 1 and 2 are shown in the accompanying Sequence Listing. The sequences are believed to represent the coding strand of the *Mycobacterium* DNA. In most instances, these sequences represent only partial sequences of the immunostimulatory peptides and, in turn, only partial sequences of *Mycobacterium tuberculosis* genes. However, each of the clones from which these sequences were derived encodes, by itself, at least one immunostimulatory T-cell epitope. As discussed in part V below, one of ordinary skill in the art will, given the information provided herein, readily be able to obtain the immunostimulatory peptides and corresponding full length *M. tuberculosis* genes using standard techniques. Accordingly, the nucleotide sequences of the present invention encompass not only those sequences presented in the sequence listings, but also the complete nucleotide sequence encoding the immunostimulatory peptides as well as the corresponding *M. tuberculosis* genes. The nucleotide abbreviations employed in the sequence listings are as follows in Table 3:

TABLE 2

Immunostimulatory AP-fusion clones (cont'd)

No.	Clone Name	INF	Fus-MW	TBport	coding	Similarity (score)
36	HinP#2-23	2,339	~43,000	~1,400	~39	Similar to sequence in *M. leprae* genomic clone MD0009-0-(B13) (354)
37	HinP#1-142	6,258	~69,000	~27,400	~741	nothing evident
38	HinP#2-4	6,567	~66,000	~24,400	~660	nothing evident
39	HinP#2-143	3,689	~65,000	~23,400	~633	Similar to sequence in *M. leprae* genomic clone B1529
40	HinP#2-145A	2,314	~64,000	~22,400	~606	nothing evident
41	HinP#2-147	7,021	65,000	23,400	~633	nothing evident
42	HinP#3-28	2,980	70,000	28,400	~768	Similar to *M. leprae* sequence in genomic clones MD0085 and sequence for *M. leprae* gp U00013 cds 27 of B1496 region
43	HinP#3-34	2,564	71,000	29,400	~795	Similar to sequence in *M. leprae* genomic clone B2168 (U00018 cds 9)
44	HinP#3-41	3,296	48,000	6,400	~1,728	Similar to antigen 85 complex protein subunit
45	HpaII#1-3	2,360	65,000	23,400	~633	Cytochrome C oxidase subunit II (156) Similar to *M. tuberculosis* sequence on clone MTCY 190.11c

Peptides designed from sequences described in this application include:

Hin P#1-200 (6 peptides)

Peptide Sequence	Peptide Name
VHLATGMAETVASFSPS	HPI1-200/2
REVVHLATGMAETVASF	HPI1-200/3
RDSREVVHLATGMAETV	HPI1-200/4
DFNRDSREVVHLATGMA	HPI1-200/5
ISAAVVTGYLRWTTPDR	HPI1-200/6
AVVFLCAAAISAAVVTG	HPI1-200/7

AciI#2-827 (14 peptides)

Peptide Sequence	Peptide Name
VTDNPAWYRLTKFFGKL	CD-2/1/96/1
AWYRLTKFFGKLFLINF	CD-2/1/96/2
KFFGKLFLINFAIGVAT	CD-2/1/96/3
FLINFAIGVATGIVQEF	CD-2/1/96/4
AIGVATGIVQEFQFGMN	CD-2/1/96/5
TGIVQEFEFGMNWSEYS	CD-2/1/96/6
EFQFGMNWSEYSRFVGD	CD-2/1/96/7
MNWSEYSRFVGDVFGAP	CD-2/1/96/8
WSEYSRFVGDVFGAPLA	CD-2/1/96/9
EYSRFVGDVFGAPLAME	CD-2/1/96/10
SRFVGDVFGAPLAMESL	CD-2/1/96/11
WIFGWNRLPRLVHLACI	CD-2/1/96/12
WNRLPRLVHLACIWIVA	CD-2/1/96/13
GRAELSSIVVLLTNNTA	CD-2/1/96/14

HinP#1-3 (2 peptides)

Peptide Sequence	Peptide Name
GKTYDAYFTDAGGITPG	HPI1-3/2
YDAYFTDAGGITPGNSV	HPI1-3/3

HinP#1-3 / HinP#1-200 combined peptides

Peptide Sequences		Peptide Name
WPQGKTYDAYFTDAGGI	(HinP#1-3)	HPI1-3/1 (combined)
ATGMAETVASFSPSEGS	(HinP#1-200)	

AciI#2-823 (1 peptide)

Peptide Sequence	Peptide Name
GWERRLRHAVSPKDPAQ	AI2-823/1

HinP#1-31 (4 peptides)

Peptide Sequence	Peptide Name
TGSGETTTAAGTTASPG	HPI1-31/1
GAAILVAGLSGCSSNKS	HPI1-31/2
AVAGAAILVAGLSGCSS	HPI1-31/3
LTVAVAGAAILVAGLSG	HPI1-31/4

These synthetic peptides were resuspended in phosphate buffered saline to be tested to confirm their ability to function as T cell epitopes using the procedure described in part IV(B)(6) above.

5. CONFIRMATION OF IMMUNOSTIMULATORY CAPACITY USING T CELLS FROM TUBERCULOSIS PATIENTS

The synthetic peptides described above, along with a number of the PhoA fusion proteins shown to be immunostimulatory in mice were tested for their ability to stimulate gamma interferon production in T-cells from tuberculin positive people using the methods described in part IV(B)(6) above. For each assay, 5×10^5 mononuclear cells were stimulated with up to 1 μg/ml *M. tuberculosis* peptide or up to 50 ng/ml Pho A fusion protein. *M. tuberculosis* filtrate proteins, Con A and PHA were employed as positive controls. An assay was run with media alone to determine background levels, and Pho A protein was employed as a negative control.

TABLE 3

Symbol	Meaning
A.....	A; adenine
C.....	C; cytosine
G.....	G; guanine
T.....	T; thymine
U.....	U; uracil
M.....	A or C
R.....	A or G
W.....	A or T/U
S.....	C or G
Y.....	C or T/U
K.....	G or T/U
V.....	A or C or G; not T/U
H.....	A or C or T/U; not G
D.....	A or G or T/U; not C
B.....	C or G or T/U; not A
N.....	(A or C or G or T/U) or (unknown or other or no base)
-.....	indeterminate*

* indicates an unreadable sequence compression.

The DNA sequences obtained were then analyzed with respect to the G+C content as a function of codon position over a window of 120 codons using the 'FRAME' computer program (Bibb, M.J.; Findlay, P.R.; and Johnson, M.W.; Gene 30: 157-166 (1984)). This program uses the bias of these nucleotides for each of the codon positions to enable the correct reading frame to be identified.

3. IDENTIFICATION OF T CELL EPITOPES IN THE IMMUNOSTIMULATORY PEPTIDES

The T-Site program, by Feller, D.C. and de la Cruz, V.F., MedImmune Inc., 19 Firstfield Rd., Gaithersburg, M.D. 20878, U.S.A., was used to predict T-cell epitopes from the determined coding sequences. It uses a series of four predictive algorithms. In particular, peptides were designed against regions indicated by the algorithm "A" motif which predicted alpha-helical periodicity (Margalit, H.; Spouge, J.L.; Cornette, J.L.; Cease, K.B.; DeLisi, C.; and Berzofsky, J.A., *J. Immunol.*, 138:2213 (1987)) and amphipathicity and those indicated by the algorithm "R" motif which identifies segments which display similarity to motifs known to be recognized by MHC class I and class II molecules (Rothbard, J.B. and Taylor, W.R., *EMBO J.* 7:93 (1988)). The other two algorithms identify classes of T-cell epitopes recognized in mice.

4. SYNTHESIS OF SYNTHETIC PEPTIDES CONTAINING T CELL EPITOPES IN IDENTIFIED IMMUNOSTIMULATORY PEPTIDES

A series of staggered peptides were designed to overlap regions indicated by the T-site analysis. These were synthesized by Chiron Mimotopes Pty. Ltd. (11055 Roselle St., San Diego, CA 92121, U.S.A.).

An alternative approach to cloning the full length ORFs corresponding to the DNA sequences provided herein is the use of the polymerase chain reaction (PCR). In particular, the inverse polymerase chain reaction (IPCR) is useful to isolate DNA sequences flanking a known sequence. Methods for amplification of flanking sequences by IPCR are described in Chapter 27 of reference no. 17 and in reference no. 23.

Accordingly, one aspect of the present invention is small oligonucleotides encompassed by the DNA sequences presented in the Sequence Listing. These small oligonucleotides are useful as hybridization probes and PCR primers that can be employed to clone the corresponding full length *Mycobacterium tuberculosis* ORFs. In preferred embodiments, these oligonucleotides will comprise at least 15 contiguous nucleotides of a DNA sequence set forth in the Sequence Listing, and in more preferred embodiments, such oligonucleotides will comprise at least 20 contiguous nucleotides of a DNA sequence set forth in the Sequence Listing.

One skilled in the art will appreciate that hybridization probes and PCR primers are not required to exactly match the target gene sequence to which they anneal. Therefore, in another embodiment, the oligonucleotides will comprise a sequence of at least 15 nucleotides and preferably at least 20 nucleotides, the oligonucleotide sequence being substantially similar to a DNA sequence set forth in the Sequence Listing. Preferably, such oligonucleotides will share at least about 75%-90% sequence identity with a DNA sequence set forth in the Sequence Listing and more preferably the shared sequence identity will be greater than 90%.

B. EXAMPLE - CLONING OF THE FULL LENGTH ORF CORRESPONDING TO CLONE HinP #2-92

Using the techniques described below, the full length gene corresponding to the clone HinP #2-92 was obtained. This gene, herein termed *mtb2-92* includes an open-reading frame of 1089 bp (identified based on the G+C content relating to codon position). The alternative 'GTG' start codon was used, and this was preceded (8 bps upstream) by a Shine-Dalgarno motif. The gene *mtb2-92* encoded a protein (termed MTB2-92) containing 363 amino acid residues with a predicted molecular weight of 40,436.4 Da.

Sequence homology comparisons of the predicted amino acid sequence of MTB2-92 with known proteins in the database indicated similarity to the cytochrome c oxidase subunit II of many different organisms. This integral membrane protein is part of the electron transport chain, subunits I and II forming the functional core of the enzyme complex.

1. CLONING THE FULL LENGTH GENE CORRESPONDING TO HinP #2-92

The plasmid pHin2-92 was restricted with either *Bam*H1 or *Eco*RI and then subcloned into the vector M13. The insert DNA fragments were sequenced under the direction of M13 universal sequencing primers (Yanisch-Perron, C. *et al.*, 1985) using the AmpliCycle thermal sequencing kit (Perkin Elmer, Applied Biosystems Division, 850 Lincoln Centre Drive, Foster City, CA 94404, U.S.A.). The 5'-partial MTB2-92 DNA sequence was aligned using a GeneWorks (Intelligenetics, Mountain View, CA, U.S.A.) program. Based on the sequence data obtained, two oligomers were synthesized. These oligonucleotides (5'CCCAGCTTGTGATACAGGAGG3' 5'GGCCTCAGCGCGGCTCCGGAGG3') represented sequences upstream and downstream, over an 0.8 kb distance, of the sequence encoding the partial MTB2-92 protein in the alkaline phosphatase fusion.

A *Mycobacterium tuberculosis* genomic cosmid DNA library was screened using PCR (Sambrook, J. *et al.*, 1989) in order to obtain the full-length gene encoding the MTB2-92 protein. Two hundred and ninety-four bacterial colonies containing the cosmid library were pooled into 10 groups in 100 µl distilled water aliquots and boiled for 5 min. The samples were spun in a microfuge at maximal speed for 5 min. The supernatants were decanted and stored on ice prior to PCR analysis.

The 100 µl-PCR reaction contained: 10 µl supernatant containing cosmid DNA, 10 µl of 10X PCR buffer, 250 µM dNTP's, 300 nM downstream and upstream primers, 1 unit *Taq* DNA polymerase.

The results, shown in Table 4 below, indicate that all of the peptides tested stimulated gamma interferon production from T-cells of a particular subject.

TABLE 4

Peptide or Pho A Fusion Protein Name	Concentration of Interferon-gamma (pg/ml)	Concentration of Interferon-gamma minus background (pg/ml)
CD-2/1/96/1	256.6	153.3
CD-2/1/96/9	187.6	84.3
CD-2/1/96/10	134.0	30.7
CD-2/1/96/11	141.6	38.3
CD-2/1/96/14	310.2	206.9
HPI1-3/2	136.3	23.0
HPI1-3/3	264.2	160.9
AciI 2-898	134.0	30.7
AciI 3-47	386.8	283.5
M. tuberculosis filtrate proteins (10 μg/ml)	256.6	153.3
M. tuberculosis filtrate proteins (5 μg/ml)	134.0	30.7
Con A (10 μg/ml)	2 839	2 735.7
PHA (1%)	10 378	10 274.7
Pho A control (10 μg/ml)	26.7	0
Background	103.3	0

V. CLONING OF FULL LENGTH *MYCOBACTERIUM TUBERCULOSIS* T-CELL EPITOPE ORFS

Most the sequences presented represent only part of a larger *M. tuberculosis* ORF. If desired, the full length *M. tuberculosis* ORFs that include these provided nucleotide sequences can be readily obtained by one of ordinary skill in the art, based on the sequence data provided herein.

A. GENERAL METHODOLOGIES

Methods for obtaining full length genes based on partial sequence information are standard in the art and are particularly simple for prokaryotic genomes. By way of example, the full length ORFs corresponding to the DNA sequences presented herein may be obtained by creating a library of *Mycobacterium tuberculosis* DNA in a plasmid, bacteriophage or phagemid vector and screening this library with a hybridization probe using standard colony hybridization techniques. The hybridization probe consists of an oligonucleotide derived from a DNA sequence according to the present invention labelled with a suitable marker to enable detection of hybridizing clones. Suitable markers include radionuclides, such as P-32 and non-radioactive markers, such as biotin. Methods for constructing suitable libraries, production and labelling of oligonucleotide probes and colony hybridization are standard laboratory procedures and are described in standard laboratory manuals such as in reference nos. 15 and 16.

Having identified a clone that hybridizes with the oligonucleotide, the clone is identified and sequenced using standard methods such as described in Chapter 13 of reference no. 15. Determination of the translation initiation point of the DNA sequence enables the ORF to be located.

At each stage of the protein purification, a sample was analysed by SDS polyacylamide gel electrophoresis (Laemmli, U.S. (1970) *Nature (London)*, 227:680-685) (see Fig. 2).

C. CORRECTION OF SEQUENCE ERRORS

It is noted that some of the sequences presented in the Sequence Listing contain sequence ambiguities. Naturally, in order to ensure that the immunostimulatory function is maintained, one would utilize a sequence without such ambiguities. For those sequences containing ambiguities, one would therefore utilize the sequence data provided in the Sequence Listing to design primers corresponding to each terminal of the provided sequence and, using these primers in conjunction with the polymerase chain reaction, synthesize the desired DNA molecule using *M. tuberculosis* genomic DNA as a template. Standard PCR methodologies, such as those described above, may be used to accomplish this.

VI. EXPRESSION AND PURIFICATION OF THE CLONED PEPTIDES

Having provided herein DNA sequences encoding *Mycobacterium tuberculosis* peptides having an immunostimulatory activity, as well as the corresponding full length *Mycobacterium tuberculosis* genes, one of skill in the art will be able to express and purify the peptides encoded by these sequences. Methods for expressing proteins by recombinant means in compatible prokaryotic or eukaryotic host cells are well known in the art and are discussed, for example, in reference nos. 15 and 16. Peptides expressed by the nucleotide sequences disclosed herein are useful for preparing vaccines effective against *M. tuberculosis* infection, for use in diagnostic assays and for raising antibodies that specifically recognize *M. tuberculosis* proteins. One method of purifying the peptides is that presented in part V(B) above.

The most commonly used prokaryotic host cells for expressing prokaryotic peptides are strains of Escherichia coli, although other prokaryotes, such as Bacillus subtilis *Streptomyces* or Pseudomonas may also be used, as is well known in the art. Partial or full-length DNA sequences, encoding an immunostimulatory peptide according to the present invention, may be ligated into bacterial expression vectors. One aspect of the present invention is thus a recombinant DNA vector including a nucleic acid molecule provided by the present invention. Another aspect is a transformed cell containing such a vector.

Methods for expressing large amounts of protein from a cloned gene introduced into *Escherichia coli* (*E. coli*) may be utilized for the purification of the *Mycobacterium tuberculosis* peptides. Methods and plasmid vectors for producing fusion proteins and intact native proteins in bacteria are described in reference no. 15 (ch. 17). Such fusion proteins may be made in large amounts, are relatively simple to purify, and can be used to produce antibodies. Native proteins can be produced in bacteria by placing a strong, regulated promoter and an efficient ribosome binding site upstream of the cloned gene. If low levels of protein are produced, additional steps may be taken to increase protein production; if high levels of protein are produced, purification is relatively easy.

Often, proteins expressed at high levels are found in insoluble inclusion bodies. Methods for extracting proteins from these aggregates are described in ch. 17 of reference no. 15. Vector systems suitable for the expression of *lacZ* fusion genes include the pUR series of vectors (24), pEX1-3 (25) and pMR100 (26). Vectors suitable for the production of intact native proteins include pKC30 (27), pKK177-3 (28) and pET-3 (29). Fusion proteins may be isolated from protein gels, lyophilized, ground into a powder and used as antigen preparations.

Mammalian or other eukaryotic host cells, such as those of yeast, filamentous fungi, plant, insect, amphibian or avian species, may also be used for protein expression, as is well known in the art. Examples of commonly used mammalian host cell lines are VERO and HeLa cells, Chinese hamster ovary (CHO) cells, and WI38, BHK, and COS cell lines, although it will be appreciated by the skilled practitioner that other prokaryotic and eukaryotic cells and cell lines may be appropriate for a variety of purposes, e.g., to provide higher expression, desirable glycosylation patterns, or other features.

The reactions were heated at 95°C for 2 min and then 40 cycles of DNA synthesis were performed (95°C for 30 s, 65°C for 1 min, 72°C for 2 min). The PCR products were loaded into a 1% agarose gel in TAE buffer (Sambrook, J. *et al.*, 1989) for analysis.

The supernatant, which produced 800 bp PCR products, was then further divided into 10 samples and the PCR reactions were performed again. The colony which had resulted in the correctly sized PCR product was then picked. The cosmid DNA from the positive clone (pG3) was prepared using the Wizard Mini-Prep Kit (Promega Corp, Madison, WI, U.S.A.). The cosmid DNA was further sequenced using specific oligonucleotide primers. The deduced amino acid sequence encoded by the MTB2-92 protein is shown in Fig. 1.

2. EXPRESSION OF THE FULL LENGTH GENE

To conveniently purify the recombinant protein, a histidine tag coding sequence was engineered immediately upstream of the start codon of *mtb2-92* using PCR. Two unique restriction enzyme sites for *Xba*I and *Hind*III were added to both ends of the PCR product for convenient subcloning. Two oligomers were used to direct the PCR reaction: (5'TCTAGACACCACCACCACCACCACGTGACACCTCGCGGGCCAGGTC3' and 5'AAGCTTCGCCATGCCGCCGGTAAGCGCC3')

The 100 μl PCR reaction contained: 1 μg pG3 template DNA, 250 μM dNTP's, 300 nM of each primer, 10 μl of 10X PCR buffer, 1 unit *Taq* DNA polymerase. The PCR DNA synthesis cycle was performed as above.

The 1.4 kb PCR products were purified and ligated into the cloning vector pGEM-T (Promega). Inserts were removed by digestion using both the *Xba*I and *Hind*III and the 1.4 kb fragment was directionally subcloned into the *Xba*I and *Hind*III sites of pMAL-c2 vector (New England Bio-Labs Ltd., 3397 American Drive, Unit 12, Mississauga, Ontario, L4V 1T8, Canada). The gene encoding MTB2-92 was fused, in frame, downstream of the maltose binding protein (MBP). This expression vector was named pMAL-MTB2-92.

3. PURIFICATION OF THE ENCODED PROTEIN

The plasmid pMAL-MTB2-92 was transformed into competent *E. coli* JM109 cells and a 1 litre culture was grown up in LB broth at 37°C to an OD_{550} of 0.5 to 0.6. The expression of the gene was induced by the addition of IPTG (0.5 mM) to the culture medium, after which the culture was grown for another 3 hours at 37°C with vigorous shaking. Cultures were spun in the centrifuge at 10,000 g for 30 min and the cell pellet was harvested. This was re-suspended in 50 ml of 20 mM Tris-HCl, pH 7.2, 200 mM NaCl, 1 mM EDTA supplemented with 10 mM β mercaptoethanol and stored at —20°C.

The frozen bacterial suspension was thawed in cold water (0°C), placed in an ice bath, and sonicated. The resulting cell lysate was then centrifuged at 10,000 g and 4°C for 30 min, the supernatant retained, diluted with 5 volumes of buffer A and applied to an amylose-resin column (New England Bio-Labs Ltd., 3397 American Drive, Unit 12, Mississauga, Ontario, L4V 1T8, Canada) which had been pre-equilibrated with buffer A. The column was then washed with buffer A until the eluate reached an A_{280} of 0.001 at which point, the bound MBP-MTB2-92 fusion protein was eluted with buffer A containing 10 mM maltose. The protein purified by the amylose-resin affinity column was about 84 kDa which corresponded to the expected size of the fusion protein (MBP: 42 kDa, MTB2-92 plus the histidine tag: 42 kDa).

The eluted MBP-MTB2-92 fusion protein was then cleaved with factor Xa to remove the MBP from the MTB2-92 protein. One ml of fusion protein (1 mg/ml) was mixed with 100 μl of factor Xa (200 μg/ml) and kept at room temperature overnight. The mixture was diluted with 10 ml of buffer B (5 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9, 6 M urea) and urea was added to the sample to a final concentration of 6 M urea. The sample was loaded onto the Ni-NTA column (QIAGEN, 9600 De Soto Ave., Chatsworth, CA 91311, U.S.A.) pre-equilibrated with buffer B. The column was washed with 10 volumes of buffer B and 6 volumes of buffer C (60 mM imidazole , 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9, 6 M urea). The bound protein was eluted with 6 volumes of buffer D (1 M imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9, 6 M urea).

TABLE 5

The Genetic Code

First Position (5' end)	Second Position				Third Position (3' end)
	T	C	A	G	
T	Phe	Ser	Tyr	Cys	T
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop (och)	Stop	A
	Leu	Ser	Stop (amb)	Trp	G
C	Leu	Pro	His	Arg	T
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	T
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	T
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val (Met)	Ala	Glu	Gly	G

"Stop (och)" stands for the ocre termination triplet, and "Stop (amb)" for the amber. ATG is the most common initiator codon; GTG usually codes for valine, but it can also code for methionine to initiate an mRNA chain.

VII. SEQUENCE VARIANTS

It will be apparent to one skilled in the art that the immunostimulatory activity of the peptides encoded by the DNA sequences disclosed herein lies not in the precise nucleotide sequence of the DNA sequences, but rather in the epitopes inherent in the amino acid sequences encoded by the DNA sequences. It will therefore also be apparent that it is possible to recreate the immunostimulatory activity of one of these peptides by recreating the epitope, without necessarily recreating the exact DNA sequence. This could be achieved either by directly synthesizing the peptide (thereby circumventing the need to use the DNA sequences) or, alternatively, by designing a nucleic acid sequence that encodes for the epitope, but which differs, by reason of the redundancy of the genetic code, from the sequences disclosed herein.

Accordingly, the degeneracy of the genetic code further widens the scope of the present invention as it enables major variations in the nucleotide sequence of a DNA molecule while maintaining the amino acid sequence of the encoded protein. The genetic code and variations in nucleotide codons for particular amino acids is presented in Tables 5 and 6. Based upon the degeneracy of the genetic code, variant DNA molecules may be derived from the DNA sequences disclosed herein using standard DNA mutagenesis techniques, or by synthesis of DNA sequences.

TABLE 7

Original Residue	Conservative Substitutions
Ala	ser
Arg	lys
Asn	gln, his
Asp	glu
Cys	ser
Gln	asn
Glu	asp
Gly	pro
His	asn; gln
Ile	leu, val
Leu	ile; val
Lys	arg; gln; glu
Met	leu; ile
Phe	met; leu; tyr
Ser	thr
Thr	ser
Trp	tyr
Tyr	trp; phe
Val	ile; leu

Substantial changes in immunological identity are made by selecting substitutions that are less conservative than those in Table 7, i.e., selecting residues that differ more significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in protein properties will be those in which (a) a hydrophilic residue, e.g., seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g., leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histadyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine. However, such variants must retain the ability to stimulate INF-γ production.

VIII. USE OF CLONED *MYCOBACTERIUM* SEQUENCES TO PRODUCE VACCINES

The purified peptides encoded by the nucleotide sequences of the present invention may be used directly as immunogens for vaccination. The conventional tuberculosis vaccine is the BCG (bacille Calmette-Guerin) vaccine, which is a live vaccine comprising attenuated *Mycobacterium bovis* bacteria. However, the use of this vaccine in a number of countries, including the U.S., has been limited because administration of the vaccine interferes with the use of the tuberculin skin test to detect infected individuals (see Cecil Textbook of Medicine (Ref. 33), pages 1733-1742 and section VIII (2) below).

The present invention provides a possible solution to the problems inherent in the use of the BCG vaccine in conjunction with the tuberculin skin test. The solution proposed is based upon the use of one or more of the immunostimulatory *M. tuberculosis* peptides disclosed herein as a vaccine and one or more different immunostimulatory *M. tuberculosis* peptides disclosed herein in the tuberculosis skin test (see section IX (2) below). If the immune system is primed with such a vaccine, it will be able to resist an infection by *M.*

TABLE 6

The Degeneracy of the Genetic Code

Number of Synonymous Codons	Amino Acid	Total Number of Codons
6	Leu, Ser, Arg	18
4	Gly, Pro, Ala, Val, Thr	20
3	Ile	3
2	Phe, Tyr, Cys, His, Gln, Glu, Asn, Asp, Lys	18
1	Met, Trp	2
Total number of codons for amino acids		61
Number of codons for termination		3
Total number of codons in genetic code		64

Additionally, standard mutagenesis techniques may be used to produce peptides which vary in amino acid sequence from the peptides encoded by the DNA molecules disclosed herein. However, such peptides will retain the essential characteristic of the peptides encoded by the DNA molecules disclosed herein, i.e. the ability to stimulate INF-γ production. This characteristic can readily be determined by the assay technique described above. Such variant peptides include those with variations in amino acid sequence including minor deletions, additions and substitutions.

While the site for introducing an amino acid sequence variation is predetermined, the mutation *per se* need not be predetermined. For example, in order to optimize the performance of a mutation at a given site, random mutagenesis may be conducted at the target codon or region and the expressed protein variants screened for the optimal combination of desired activity. Techniques for making substitution mutations at predetermined sites in DNA having a known sequence as described above are well known.

In order to maintain the functional epitope, preferred peptide variants will differ by only a small number of amino acids from the peptides encoded by the DNA sequences disclosed herein. Preferably, such variants will be amino acid substitutions of single residues. Substitutional variants are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the following Table 7 when it is desired to finely modulate the characteristics of the protein. Table 7 shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative substitutions. As noted, all such peptide variants are tested to confirm that they retain the ability to stimulate INF-γ production.

microorganism as a vaccine. As described in International Patent Application WO 95/01441, *Mycobacterium bovis* BCG may be employed for this purpose, although this approach would destroy the advantage outlined above to be gained from using separate classes of the peptides as vaccines and in the skin test. As disclosed in WO 95/01441, an immunostimulatory peptide of *M. tuberculosis* can be expressed in the BCG bacterium by transforming the BCG bacterium with a nucleotide sequence encoding the *M. tuberculosis* peptide. Thereafter, the BCG bacteria can be administered in the same manner as a conventional BCG vaccine. In particular embodiments, multiple copies of the *M. tuberculosis* sequence are transformed into the BCG bacteria to enhance the amount of *M. tuberculosis* peptide produced in the vaccine strain.

IX. USE OF CLONED *MYCOBACTERIUM* SEQUENCES IN DIAGNOSTIC ASSAYS

Another aspect of the present invention is a composition for diagnosing tuberculosis infection wherein the composition includes peptides encoded by the nucleotide sequences of the present invention. The invention also encompasses methods and compositions for detecting the presence of anti-tuberculosis antibodies, tuberculosis peptides and tuberculosis nucleic acid sequences in body samples. Three examples typify the various techniques that may be used to diagnose tuberculosis infection using the present invention: an in vitro ELISA assay, an in vivo skin test assay and a nucleic acid amplification assay.

A. IN VITRO ELISA ASSAY

One aspect of the invention is an ELISA that detects anti-tuberculosis mycobacterial antibodies in a medical specimen. An immunostimulatory peptide encoded by a nucleotide sequence of the present invention is employed as an antigen and is preferably bound to a solid matrix such as a crosslinked dextran such as SEPHADEX (Pharmacia, Piscataway, NJ), agarose, polystyrene, or the wells of a microtiter plate. The polypeptide is admixed with the specimen, such as human sputum, and the admixture is incubated for a sufficient time to allow antimycobacterial antibodies present in the sample to immunoreact with the polypeptide. The presence of the immunopositive immunoreaction is then determined using an ELISA assay.

In a preferred embodiment, the solid support to which the polypeptide is attached is the wall of a microtiter assay plate. After attachment of the polypeptide, any nonspecific binding sites on the microtiter well walls are blocked with a protein such as bovine serum albumin. Excess bovine serum albumin is removed by rinsing and the medical specimen is admixed with the polypeptide in the microtiter wells. After a sufficient incubation time, the microtiter wells are rinsed to remove excess sample and then a solution of a second antibody, capable of detecting human antibodies is added to the wells. This second antibody is typically linked to an enzyme such as peroxidase, alkaline phosphatase or glucose oxidase. For example, the second antibody may be a peroxidase-labeled goat anti-human antibody. After further incubation, excess amounts of the second antibody are removed by rinsing and a solution containing a substrate for the enzyme label (such as hydrogen peroxide for the peroxidase enzyme) and a color-forming dye precursor, such as o-phenylenediamine is added. The combination of mycobacterium peptide (bound to the wall of the well), the human antimycobacterial antibodies (from the specimen), the enzyme-conjugated anti-human antibody and the color substrate will produce a color than can be read using an instrument that determines optical density, such as a spectrophotometer. These readings can be compared to a control incubated with water in place of the human body sample, or, preferably, a human body sample known to be free of antimycobacterial antibodies. Positive readings indicate the presence of anti-mycobacterial antibodies in the specimen, which in turn indicate a prior exposure of the patient to tuberculosis.

B. SKIN TEST ASSAY

Alternatively, the presence of tuberculosis antibodies in a patient's body may be detected using an improved form of the tuberculin skin test, employing immunostimulatory peptides of the present invention. Conventionally, this test produces a positive result to one of the following conditions: the current presence of *M. tuberculosis* in the patient's body; past exposure of the patient to *M. tuberculosis*; and prior BCG vaccination. As

tuberculosis. However, exposure to the vaccine peptides alone will not induce an immune response to those peptides that are reserved for use in the tuberculin skin test. Thus, the present invention would allow the clinician to distinguish between a vaccinated individual and an infected individual.

Methods for using purified peptides as vaccines are well known in the art and are described in the following publications: Pal and Horwitz (1992) (reference no. 8) (describing immunization with extra-cellular proteins of *Mycobacterium tuberculosis*); Yang et al. (1991) (reference no. 30) (vaccination with synthetic peptides corresponding to the amino acid sequence of a surface glycoprotein from *Leishmania major*); Andersen (1994) (reference no. 9) (vaccination using short-term culture filtrate containing proteins secreted by *Mycobacterium tuberculosis*); and Jardim et al. (1990) (reference no. 10) (vaccination with synthetic T-cell epitopes derived from *Leishmania* parasite). Methods for preparing vaccines which contain immunogenic peptide sequences are also disclosed in U.S. Patent Nos. 4,608,251, 4,601,903, 4,599,231, 4,5995230, 4,596,792 and 4,578,770. The formulation of peptide-based vaccines employing *M. tuberculosis* peptides is also discussed extensively in International Patent application WO 95/01441.

As is well known in the art, adjuvants such as Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA) may be used in formulations of purified peptides as vaccines. Accordingly, one embodiment of the present invention is a vaccine comprising one or more immunostimulatory *M. tuberculosis* peptides encoded by genes including a sequence shown in the attached sequence listing, together with a pharmaceutically acceptable adjuvant.

Additionally, the vaccines may be formulated using a peptide according to the present invention together with a pharmaceutically acceptable excipient such as water, saline, dextrose and glycerol. The vaccines may also include auxillary substances such as emulsifying agents and pH buffers.

It will be appreciated by one of skill in the art that vaccines formulated as described above may be administered in a number of ways including subcutaneous, intra-muscular and intra-venous injection. Doses of the vaccine administered will vary depending on the antigenicity of the particular peptide or peptide combination employed in the vaccine, and characteristics of the animal or human patient to be vaccinated. While the determination of individual doses will be within the skill of the administering physician, it is anticipated that doses of between 1 microgram and 1 milligram will be employed.

As with many vaccines, the vaccines of the present invention may routinely be administered several times over the course of a number of weeks to ensure that an effective immune response is triggered. As described in International Patent Application WO 95/01441, up to six doses of the vaccine may be administered over a course of several weeks, but more typically between one and four doses are administered. Where such multiple doses are administered, they will normally be administered at from two to twelve week intervals, more usually from three to five week intervals. Periodic boosters at intervals of 1-5 years, usually three years, will be desirable to maintain the desired levels of protective immunity.

As described in WO 95/01441, the course of the immunization may be followed by *in vitro* proliferation assays of PBL (peripheral blood lymphocytes) co-cultured with ESAT6 or ST-CF, and especially by measuring the levels of IFN-γ released from the primed lymphocytes. The assays are well known and are widely described in the literature, including in U.S. Patent Nos. 3,791,932; 4,174,384 and 3,949,064.

To ensure an effective immune response against tuberculosis infection, vaccines according to the present invention may be formulated with more than one immunostimulatory peptide encoded by the nucleotide sequences disclosed herein. In such cases, the amount of each purified peptide incorporated into the vaccine will be adjusted accordingly.

Alternatively, multiple immunostimulatory peptides may also be administered by expressing the nucleic acids encoding the peptides in a nonpathogenic microorganism, and using this transformed nonpathogenic

methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Harlow and Lane (1988).

B. ANTIBODIES RAISED AGAINST SYNTHETIC PEPTIDES.

An alternative approach to raising antibodies against the *M. tuberculosis* peptides is to use synthetic peptides synthesized on a commercially available peptide synthesizer based upon the amino acid sequence of the peptides predicted from nucleotide sequence data.

In a preferred embodiment of the present invention, monoclonal antibodies that recognize a specific *M. tuberculosis* peptide are produced. Optimally, monoclonal antibodies will be specific to each peptide, i.e. such antibodies recognize and bind one *M. tuberculosis* peptide and do not substantially recognize or bind to other proteins, including those found in healthy human cells.

The determination that an antibody specifically detects a particular *M. tuberculosis* peptide is made by any one of a number of standard immunoassay methods; for instance, the Western blotting technique (Sambrook et al., 1989). To determine that a given antibody preparation (such as one produced in a mouse) specifically detects one *M. tuberculosis* peptide by Western blotting, total cellular protein is extracted from a sample of human sputum from a healthy patient and from sputum from a patient suffering from tuberculosis. As a positive control, total cellular protein is also extracted from *M. tuberculosis* cells grown in vitro. These protein preparations are then electrophoresed on a sodium dodecyl sulfate-polyacrylamide gel. Thereafter, the proteins are transferred to a membrane (for example, nitrocellulose) by Western blotting, and the antibody preparation is incubated with the membrane. After washing the membrane to remove non-specifically bound antibodies, the presence of specifically bound antibodies is detected by the use of an anti-mouse antibody conjugated to an enzyme such as alkaline phosphatase; application of the substrate 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium results in the production of a dense blue compound by immuno-localized alkaline phosphatase. Antibodies which specifically detect the *M. tuberculosis* protein will, by this technique, be shown to bind to the *M. tuberculosis*-extracted sample at a particular protein band (which will be localized at a given position on the gel determined by its molecular weight) and to the proteins extracted from the sputum from the tuberculosis patient. No significant binding will be detected to proteins from the healthy patient sputum. Non-specific binding of the antibody to other proteins may occur and may be detectable as a weak signal on the Western blot. The non-specific nature of this binding will be recognized by one skilled in the art by the weak signal obtained on the Western blot relative to the strong primary signal arising from the specific antibody-tuberculosis protein binding. Preferably, no antibody would be found to bind to proteins extracted from healthy donor sputum.

Antibodies that specifically recognize a *M. tuberculosis* peptide encoded by the nucleotide sequences disclosed herein are useful in diagnosing the presence of tuberculosis antigens in patients.

All publications and published patent documents cited in this specification are incorporated herein by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

noted above, if one group of immunostimulatory peptides is reserved for use in vaccine preparations, and another group reserved for use in the improved skin test, then the skin test will not produce a positive response in individuals whose only exposure to tuberculosis antigens was via the vaccine. Accordingly, the improved skin test would be able to properly distinguish between infected individuals and vaccinated individuals.

The tuberculin skin test consists of an injection of proteins from *M. tuberculosis* that are injected intradermally. The test is described in detail in Cecil Textbook of Medicine (Ref. 33), pages 1733-1742. If the subject has reactive T-cells to the injected protein, the cells will migrate to the site of injection and cause a local inflammation. This inflammation, which is generally known as delayed type hypersensitivity (DTH) is indicative of *M. tuberculosis* antibodies in the patient's blood stream. Purified immunostimulatory peptides according to the present invention may be employed in the tuberculin skin test using the methods described in reference 33.

C. NUCLEIC ACID AMPLIFICATION

One aspect of the invention includes nucleic acid primers and probes derived from the sequences set forth in the attached sequence listing, as well as primers and probes derived from the full length genes that can be obtained using these sequences. These primers and probes can be used to detect the presence of *M. tuberculosis* nucleic acids in body samples and thus to diagnose infection. Methods for making primers and probes based on these sequences are well known and are described in section V above.

The detection of specific pathogen nucleic acid sequences in human body samples by polymerase chain reaction amplification (PCR) is discussed in detail in reference 17, in particular, part four of that reference. To detect *M. tuberculosis* sequences, primers based on the sequences disclosed herein would be synthesized, such that PCR amplification of a sample containing *M. tuberculosis* DNA would result in an amplified fragment of a predicted size. If necessary, the presence of this fragment following amplification of the sample nucleic acid could be detected by dot blot analysis (see chapter 48 of reference 17). PCR amplification employing primers based on the sequences disclosed herein may also be employed to quantify the amounts of *M. tuberculosis* nucleic acid present in a particular sample (see chapters 8 and 9 of reference 17). Reverse-transcription PCR using these primers may also be utilized to detect the presence of *M. tuberculosis* RNA, indicative of an active infection.

Alternatively, probes based on the nucleic acid sequences described herein may be labelled with suitable labels (such a P^{32} or biotin) and used in hybridization assays to detect the presence of *M. tuberculosis* nucleic acid in provided samples.

X. USE OF CLONED *MYCOBACTERIUM* SEQUENCES TO RAISE ANTIBODIES

Monoclonal antibodies may be produced to the purified *M. tuberculosis* peptides for diagnostic purposes. Substantially pure *M. tuberculosis* peptide suitable for use as an immunogen is isolated from the transfected or transformed cells as described above. The concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few milligrams per milliliter. Monoclonal antibody to the protein can then be prepared as follows:

A. MONOCLONAL ANTIBODY PRODUCTION BY HYBRIDOMA FUSION.

Monoclonal antibody to epitopes of the *M. tuberculosis* peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler and Milstein (1975) or derivative methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected purified protein over a period of a few weeks. The mouse is then sacrificed, and the antibody-producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall (1980), and derivative

24. Ruther et al. (1983). EMBO J. 2:1791.

25. Stanley and Luzio (1984). EMBO J. 3:1429.

26. Gray et al. (1982). Proc. Natl. Acad. Sci. USA 79:6598.

27. Shimatake and Rosenberg (1981). Nature 292:128.

28. Amann and Brosius (1985). Gene 40:183.

29. Studiar and Moffatt (1986). J. Mol. Biol. 189:113.

30. Yang et al. (1991). Identification and Characterization of Host-Protective T-Cell Epitopes of a Major Surface Glycoprotein (gp63) from *Leishmania major*. Immunology 72:3-9.

31. Mdluli et al. (1995). New vectors for the in vitro generation of alkaline phosphatase fusions to proteins encoded by G+C-rich DNA. Gene 155:133-134.

32. Lim et al. (1995). Identification of *Mycobacterium tuberculosis* DNA Sequences Encoding Exported Proteins by Using *phoA* Gene Fusions. J. Bact. 177:59-65).

33. Cecil Textbook of Medicine. (1992. 19th edition). Wyngaarden et al. eds. W.B. Saunders, Philadelphia, PA.

34. Hubbard et al. (1992). Immunization of mice with mycobacterial culture filtrate culture proteins. Clin. exp. Immunol. 87: 94-98.

XI. REFERENCES

1. Skamene, E. (1989). Genetic control of susceptibility to Mycobacterial infections. Ref. Infect. Dis. 11:S394-S399.

2. Kaufmann, S.H.E. (1991). Role of T-Cell Subsets in Bacterial Infections. Current Opinion in Immunology 3:465-470.

3. Orme, I.M., et al. (1992). T Lymphocytes Mediating Protection and Cellular Cytolysis During the Course of Mycobacterium-Tuberculosis Infection - Evidence for Different Kinetics and Recognition of a Wide Spectrum of Protein Antigens. Journal of Immunology 148:189-196.

4. Daugelat, S., et al. (1992). Secreted Antigens of Mycobacterium tuberculosis: characterization with T Lymphocytes from Patients and Contacts after Two-Dimensional Separation. J. Infect. Dis. 166:186-190.

5. Barnes et al. (1989). Characterization of T Cell Antigens Associated with the Cell Wall Protein-Peptidoglycan Complex of *Mycobacterium tuberculosis*. J. Immunol. 143:2656-2662.

6. Collins et al. (1988). Biological activity of protein antigens isolated from *Mycobacterium tuberculosis* culture filtrate. Infect. Immun. 56:1260-1266.

7. Lamb et al. (1989). Identification of Mycobacterial Antigens Recognized by T Lymphocytes. Rev. Infect. Dis. 11:S443-S447.

8. Pal, P.G., et al. (1992). Immunization with Extracellular Proteins of Mycobacterium tuberculosis Induces Cell-Mediated Immune Responses and Substantial Protective Immunity in a Guinea Pig Model of Pulmonary Tuberculosis. Infect. Immun. 60:4781-4792.

9. Andersen (1994). Infection & Immunity 62:2536.

10. Jardim et al. (1990). Immunoprotective *Leishmania major* Synthetic T Cell Epitopes. J. Exp. Med. 172:645-648.

11. Orme et al. (1993). Cytokine Secretion by CD4 T Lymphocytes Acquired in Response to *Mycobacterium tuberculosis* Infection. J. Immunology 151:518-525.

12. Boesen et al. (1995). Human T-Cell Responses to Secreted Antigen Fractions of *Mycobacterium tuberculosis*. Infection and Immunity 63:1491-1497.

13. Mougneau et al. (1995). Expression Cloning of a Protective *Leishmania* Antigen. Science 268:536-566.

14. Yang et al. (1990). Oral *Salmonella typhimurium* (AroA$^-$) Vaccine Expressing a Major Leishmanial Surface Protein (gp63) Preferentially Induces T Helper 1 Cells and Protective Immunity Against Leishmaniasis. J. Immunology 145:2281-2285.

15. Sambrook et al. (1989). Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, ed. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.

16. Ausubel et al., (1987). Current Protocols in Molecular Biology, ed. Greene Publishing and Wiley-Interscience: New York (with periodic updates).

17. Innis et al., (1990). PCR Protocols: A Guide to Methods and Applications, Academic Press: San Diego.

18. Kanehisa (1984). Nuc. Acids Res. 12:203-213, 1984.

19. Wetmur et al. (1968). J. Mol. Biol. 31:349-370.

20. Beaucage et al. (1981) Tetra. Letts. 22:1859-1862.

21. Matteucci et al. (1981). J. Am. Chem. Soc. 103:3185.

22. Jacobs et al. (1991) METHODS IN ENZYMOLOGY 204:537-555.

23. Earp et al. (1990). Nucleic Acids Research 18:3721-3729.

(A) LENGTH: 265
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#1-62
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 1

ACGCGGACCT CGAAGTTCAT CATCGAGTGA TACGTGCCAC ACATCTCGGC 50
GCAGTGGCCC ACGAATGCAN CCGGTCTTGG TGATTTCNTC GATCTGGAAG 100
ACGTTGACCG ARTTGTTTGC CACCGGGTTA GGCATCACGT CACGCTTGAA 150
CAAGAACTCC GGCACCCAGA ATGCGTGTGT CACATCGGCT GAGGCCATTT 200
GGAATTCGAT ACGCTTGCCG GACGGCAGCA CCAGCACCGG AATTTCGGTG 250
CTGTGCAACG TCTCG 265

(2) INFORMATION FOR SEQ ID NO: 2
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 484
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#1-152
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 2

CTGGTACGAC GCCGGCAAGG ACTACGGACG AGGTGGCACA GAATTCAATG 50
CGGCGCTCAT CGGAACCGAC GTGCCCGACG NCGTTTGCTC GACGACGATG 100
GTGNTTCCAN TTCGCCTNAN CGGTGTNCTG ACTGCCNTTG ACGACCTGNT 150
CGGCCARGTT GGGNTGGACA CAACGGATTA CGTCGATTCG CTGCTGGCCG 200
ACTATGAGTT CAACGGCCGC CATTACGCTG TGCCGTATGC TCGCTCGACG 250
CCGCTGTTCT ACTACAACAA GGCGGCGTGG CAACAGGCCG GCCTACCCGA 300
CCGCGGACCG CAATCCTGGT CAGAGTTCGA CGAGTGGGGT CCGGAGTTAC 350
AGCGCGTGGT CGNCGCCGGT CGATCGGCGC ACGGCTGCGT AACGCCGACC 400
TCATCTCGTG GACGTTTCAG GGACCGAACT GGGCATNCGG CGGTGCCTAC 450
TCCGACAAGT GGACATTGAC ATTGACCGAG CCCG 484

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANTS: UNIVERSITY OF VICTORIA INNOVATION AND DEVELOPMENT CORPORATION

(ii) TITLE OF INVENTION: *MYCOBACTERIUM TUBERCULOSIS* DNA SEQUENCES ENCODING IMMUNOSTIMULATORY PEPTIDES

(iii) NUMBER OF SEQUENCES: 76

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Klarquist Sparkman Campbell Leigh & Whinston, LLP

(B) STREET: One World Trade Center, Suite 1600, 121 S.W. Salmon Street

(C) CITY: Portland

(D) STATE: OR

(E) COUNTRY: USA

(F) ZIP: 97204-2988

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Disk, 3.5-inch

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: MS DOS

(D) SOFTWARE: WordPerfect 5.1+, ASCII

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: PCT/US96/10375

(B) FILING DATE: June 14, 1996

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 06/000,254

(B) FILING DATE: 06/15/95

(viii) ATTORNEY/AGENT INFORMATION

(A) NAME: Richard J. Polley

(B) REGISTRATION NUMBER: 28,107

(C) REFERENCE/DOCKET NUMBER: 2847-45176/RJP

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (503) 226-7391

(B) TELEFAX: (503) 228-9446

(2) INFORMATION FOR SEQ ID NO: 1

(i) SEQUENCE CHARACTERISTICS:

TGCCGACGTT GCCCTCGCCG ACGTTCGCCA AGCCCAGGTT GCGGACACGC 200
CGGTGATTGT GCGTGGGGCA ATGACGGGCT GCTGGCCCGG CCGAATTCCA 250
AGGCGTCGAT CGGCACGGTG TTCCAGGACC GGGCCGCTCG CTACGGTGAC 300
CGAGTCTTCC TGAAATTCGG CGATCAGCAG CTGACCTACC GCGACCGTAA 350
CGCCACCGCC AACCGGTNNG CCGCGGTGTT GGCCNNNCGC GGCGTCGGCC 400
CCGGCGACGT CGTTGGCATC ATGTTGCGTA ACTCACCCAG CACAGTCTTG 450
GCGATGCTGG CCACGGTCAA GTGCGGCGTA TCGCCGGCAT GCTCAACTAC 500
CACCAGCGCG 510

(2) INFORMATION FOR SEQ ID NO: 5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 456

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#1-426

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 5

GCAACGGAGA GGTGGACTAT GCCGGACCGG CACCGCGAAG GGGTTGGTGC 50
CGGCCCGGGT GGTGACGGTG CACATTCTGC GCAATTCGCT GAGTTCCGGT 100
GGTGACCTTC CTGGGCGCGG AGTCTGGGCG CGCTGATGGC GGAGCGAKTG 150
TGACCGAAGG AANTCNGTTC AACATCCACG GCGTCGGGGG CGTGCTGTAT 200
CAAGCGGTCA CCGTCAGGAG ACGCCGACGG TGGTGTCGAT CGTGACGGTG 250
CTGGTGCTGA TCTACCTGAT CACCAATCTG TTGGTGGATC TGCTGTATGC 300
GGCCCTGGAC GCCGNNGATN CGCTATGGCT GAGCACACGG GGTTCTGGCT 350
CGATGCCTNG CGCGGGTTGC GCCGGCGTCC TAAANTCGTG ATCGCGCGGC 400
GCTGAKCCTG CTGATTCTTG TCGTGGCGGC GTTTCCGTCG TTGTTTACCG 450
CAGCCG 456

(2) INFORMATION FOR SEQ ID NO: 6

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 175

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(2) INFORMATION FOR SEQ ID NO: 3

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 513

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#1-239

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 3

GGCGGCCAGA CGTCGGAACT CGCGGCCAAT TGGTGTGGTG GGAACCGCGA 50

TCCTCGACGC AACCGCTTCG CGGTCTTGGC AGTGTTCGAT GCCAATCTGC 100

CGGCCGGGAC GCTGCCGGAT GCGGCCCGTT CACCGAGGCT GGTGACAAGA 150

CCTGGCGTTG TCGTTCCGGG CACTACTCCC NAGGTCGGTC AAGGCACCGT 200

CAAAGTGTTC AGGTATACCG TCGAGATCGA GAACGGTCTT GATCCCACAA 250

TGTACGGCGG TGACAANNNN ATTCGCCCAG ATGGTCGACC AGACGTTGAC 300

CAATCCCAAG GGCTGGACCC ACAATCCGCA ATTCGGCGTT CGTGCGGATC 350

GACAGCGGAA AACCCGACTT CCGGATTTCG CTGGTGTCGC CGACGACAGT 400

GCGCGGGGGN TGTGGCTACG AATTCCGGCT CGAGACGTCC TGCTACAACC 450

CGTCGTTCGG CGGCATGGAT CGCCAATCGC GGGTGTTCAT CAACGAGGCG 500

CGCTGGGTAC GCG 513

(2) INFORMATION FOR SEQ ID NO: 4

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 510

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#1-247

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 4

GTGTGCAACC AGTGTGTGTN CGTGTGCGAA CCAGTGTGTA GTGGTAACCA 50

GGACCACGTT GCAAACCAGT GTTGGAGTGC AGTGTTGCGT GCNAGTGTTG 100

CNCGTTGCAG TGTTNGNCGA GCCGAGATTG GAAGTTNCCG ACATTACCGT 150

GTTCGNCGCG CTCAAAAGGT TGACGATGGT CACGTCGCAC GTGCTGGCCG 50
AGACCAAGGT GGATTTCGGT GAAGACCTCA AAGANCTCTA CTCGNATCGT 100
CAAGGCCCTC AACGACGACC GAAAGGATTT CGTCACCTCG CTGCAGCTGT 150
TGCTGACGTT CCCATTTCCC AAC 173

(2) INFORMATION FOR SEQ ID NO: 9 :-

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 223

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 9

CCTGTTNCAA CGGTNCNTTC NCGGAACGGA CGACTTCTGA TNCGNNCTCG 50
GNCGTTCCCT CGCACCGGTC GATGGTGATC AAGGTCAGCG TCTTCGCGGT 100
GGTCATGCTG CTGGTGGCCG CCGGTCTGGT GGTGGTATTC GGGGACTTCC 150
GGTTTGGTCC CACAACCGTC TACCACGCCA CCTTCACCGA CNCGTNGCGG 200
CTGAANGCAG GCCAGAAGGT TCG 223

(2) INFORMATION FOR SEQ ID NO: 10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 120

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-272

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 10

CAACGAGATC GCACCCGTGA TTAGGAGGTG ACGGTGGCAG CGCCGACCCC 50
GTCGAATCGG ATCGAAGTAA CGCTCCGTAG ACGCCAGCTC GTCCGCGCCG 100
ATGCCGACCT GCCACCCGTG 120

(2) INFORMATION FOR SEQ ID NO: 11

(i) SEQUENCE CHARACTERISTICS:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 6

TCNCTTANYC CTTCANCTGN CATCTNTCCC AANNACCGAA NTCTGGACCT 50

ATSACGNCCA NCTNAANATG NCCCNCGACN AAGGNCNTTG NACGTTCNCT 100

GKACCACCAN CGGGTTGCAT SCCAAGCTAG NCGAACATCA NASGTTNCGC 150

GCNTACGAGC CGACCCGCCG CGGCG 175

(2) INFORMATION FOR SEQ ID NO: 7

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 231

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-23

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 7

CTTCTCGCGC CAGCCGTCCC GCTGTCCGGG ATGCGCTACC GGTCGTCAGC 50

GCCAAGACGG TGCAGCTCAA CGACGGCGGG TTGGTGCGCA CGGTGCACTT 100

GCCGGCCCCC AATGTSGCGG GGCTGCTGAG TGCGGCCGCG TGCCGCTGTT 150

GCAAANNGCG ACCACGTGGT GCCCGCCGCG ACGGCCCCGA TCGTCGAAGG 200

CATGCAGATC CAGGTGACCC GCAAATCGGA T 231

(2) INFORMATION FOR SEQ ID NO: 8

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 173

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-26

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 8

(ix) FEATURE:
(D) OTHER INFORMATION: AciI#2-511
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 13

GCGNACNCTG CGCATNGCTG CCNGTANCCC GGCGCCNAGG CATGAGNCNN 50
TAGGCCGAAA TGCCTGGTKA ANCTNGCGTG TSGTGGTTGA CCCGCNGCGT 100
SCNGGCNTAC AKGTGCATGC TGTNGATCGG CAGTGGGAGA GGTGAGCGGT 150
GCGGCGTNAA GGTGCGGAGG TTNGASNTCT GGCGGTGTCG GCGTTNGGTG 200
GCTTTGTTCC CGGCGGTCGC GGGGTGCTCC NGNATTCCGG CGACNAACNA 250
AANNCCGGGN AGSACGAYNC CCGTCGACAC CNGGCAAACG CTGAGGGCCG 300
GCACGGACCC TTCTTCCCGC AATGTGGCGG CGTCAGCGAT CANGACGGTG 350
ACCGAGCTGW ACAAGGGTGA CCGGGCTGGT CAACACCGCC AAGAAGTCGG 400
TGGGCTNCCA ATGGCNTGGC G 421

(2) INFORMATION FOR SEQ ID NO: 14
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 175
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#2-523
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 14

CCAGNCCNCC NAACNTGTYN CGNTCTCAYY TCGCCGTCGC TGCCGGTNCG 50
TGTGTGCACC ATCTGCACCG ACCCGTGKAA CYTCGATCAC GANACTGGNA 100
GAGNTCAGGC ATNAAAGCCG GAGTGGCACA GCAACGGTCG CTACTGGAAT 150
TGGCGAAGCT GGATGCTGAG CTGAC 175

(2) INFORMATION FOR SEQ ID NO: 15
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 263
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:

(A) LENGTH: 160
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#2-506
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 11

CNGGCNNCCA NCGGGTGCGC CAWGCACGGC CGGTCCGTGC GAGATCGTCN 50
CNAATGGCAN GCCGGCGCCC AAKANANNNC CGGTACCGTG CCTTCGTNGW 100
GCAWCCTNGC GACCAACCCC GAGATYGCYA CNCTACNGCC GGKACATGAC 150
CGTGGTGCGG 160

(2) INFORMATION FOR SEQ ID NO: 12
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 133
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#2-508
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 12

GACTGGNCCC GAYGYTGTGN CCGGHNCGTH GGNCGHGCHG CANTCGAYCC 50
TGGCCGTTGC TTCGGTGCCG GGTTGTTCAT CGCCTTCGAC CAGTTGTGGC 100
GCTGGAACAG CATAGTGGCG CTAGTGCTAT CGG 133

(2) INFORMATION FOR SEQ ID NO: 13
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 421
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

CCGGGATTGG CACGGTCGGC AABGCATTCG TCAGCNNTGC GCTCGAAGGT 150
CAACAAGAAT GTCGGGGTCT ACGCGGTGAA A 181

(2) INFORMATION FOR SEQ ID NO: 18

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 95

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-872

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 18

AGGTKACGGT GGCAGCGCCG ACCCCGTCGA ATCGGWTCGA AGAAYGCTCC 50
GKACACGCCA GCTGCGTCCG YGCCGATGCC GACCTGCCAC CCGTG 95

(2) INFORMATION FOR SEQ ID NO: 19

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 65

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-884d

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 19

AKCGGTCACC KACGGGCCGG CCACCGATGC GATTGTCAAC GGATTCCAAG 50
TGGTTGYGCA TGCGC 65

(2) INFORMATION FOR SEQ ID NO: 20

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 156

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(D) OTHER INFORMATION: AciI#2-639

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 15

GGGCTGGATT CGAGGCTCGT GCATGNCGTA CGACTANGGG TAGCGCCCAG 50
CTGCTCAATA CCATCGGTTG GATAACAAAG GCTGAACATG AATGGCNTGA 100
TCTCNACAAG CGTGCGGCTC CCACCGACCC CGGCGCCCCT CGAGCCTGGG 150
GSTGTCGCGA TCCTGATCGC GGCGACACTT TTCGCGACTG TCGTTGCGGG 200
GTGCGGGAAA AAACCGACCA CGGCGAGCTC CCGAGTCCCG GGTCGCCGTC 250
GCCGGAAGCC CAC 263

(2) INFORMATION FOR SEQ ID NO: 16

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 168

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-822

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 16

YGCCATGCGA AGCGCACCCC GGTCCGGAAG NCCTGCACAG TTCWNCCGTG 50
CTCGCCGCGA CGCTACTCCT CGNYTGCGGC GGTCCCAYGC AGCCAYGCAG 100
CATCACCTTG ACCTTTATCC GCAACGYGYA ATYCCAGGCC AAYGCCGAYG 150
GGATCATCGA YACCKACA 168

(2) INFORMATION FOR SEQ ID NO: 17

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 181

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-854

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 17

ACCNGTTCCC GCCGGNCTNA CNCNCGGTGC CGTTGCACCG GCCANCTGCA 50
GCCTGCCCCG ACGCCGAAGT GGTGTTCGCN CCGCGGCCGC TTCGAACCGC 100

GGANCTCGAC ATCAATKCAN CCGGAGNAGN ANGCTGACCN AACATNCGCT 200
CATCGACCGC GGATGTCNAT CGAGNACGST GCCAAGSCGC TGCAGCTGGA 250
TNCTCGAGCG CGCCATGGAG TNATRTGCGS CCGACGAATN CGTCGAGGTG 300
ACCCCGGAGA NTCGTGCGGA TSCGCRAAGT CGAGCTGGCC GGCCNGCCGC 350
CCGGGCTNMG CAGCCGGGCG CGCACCNAAG GCGCGTGGCN TAGCANACTT 400
GGCGNGCTGG CCGCGCGAGC GTANACNGCC ACTGCGAAAN TCCANGCCCG 450
GCTTTTCGCA GCCGGGTTNA CGCTCGTGGG GGTACTGGAT AGCCTGATGG 500
GCGTGCCCAG NCCCANGTCC GCCGCGTCTG TGTGACGGTC GGCGCGTTGG 550
TCGCGCTGGC GTGTATGGTG TTGGCCGGGT GCACGGTCAG CCCGCCGCCG 600
GCACCCCAGA GCASTGATAC GCCGCGCAGC ACACCG 636

(2) INFORMATION FOR SEQ ID NO: 23

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 103

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-916

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 23

CTTCCGGCGG GACAACAACA GGTCTCACCG GCGCCACACC CTGACACCTG 50
ATCGCGTCTG CCGATCCCGG TCGGAGCACC CGGGTTCCAC CGCTGTGCCC 100
CCC 103

(2) INFORMATION FOR SEQ ID NO: 24

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 207

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-1014

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 24

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-8841

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 20:

TCTTCTACAA GGACGCCTTC GCCAAGCACC AGGAGCTGTT CGACGACTTG 50

GNCGTCAACG TCAACAATGG CTTGTCCGAT CTGTACRAGC AAGWTCGAGT 100

CGCTGCCGNB CGCAACGCGA CGAGATCATC GAGGACCTAC ACCGTTGCCA 150

CGAACA 156

(2) INFORMATION FOR SEQ ID NO: 21

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 123

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-8941

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 21

ATNCCGTTCC ACTNCCGCGG CAGCAGCTGG NTTTGCGCAC ACGGTGACCC 50

AGTGGCGNTT GGTGGGGCCT CGCTGACGGC GAGTNTGGNC GAGCGTCCTC 100

GGTCGGTGNC CTNTCNTCCC GCC 123

(2) INFORMATION FOR SEQ ID NO: 22

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 636

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-898

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 22

CGGTCWHKCA ANTTGATGBC NGCGCGCAAG GCCGNCATGG TNGAGATGCC 50

AACCACACCA CCGGCTGGNT CCGCATGGAC TTCGTGNTTS CCAGTCGCNG 100

CCTGATTGGG TGNCGCACCG ACNNCCTNCA CCGAGACCSG TGGCTCNSGA 150

CGGTCCTGGC TCTCGGGCTG CTGGCCNCTG CGCCCCACCC CGCACCGGGC CGGCTTC 207

(2) INFORMATION FOR SEQ ID NO: 27

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 289

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-1084

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 27

YCNAGNCKCG TNATNGCSGN CKCATNTNAC NGGANCCNGG ATTNCSTACG 50
CCACNGTGAT CGCGCTGGTN GCCGCGCTGG TGGCGCGTGT ACGTGCTCTC 100
GTCCACCGGN AANTAAGCGC ACCATCGTGG GCTACTTCAC CTCTGCTGTC 150
GGGCTCTATC CCGGTGACCA GGTCCGCGTC CTGGGCGTCC NGGTGGGTGA 200
GATCGACATG ATCGAGCCGC GGTCGTCCGA CGTSAAGATC ACTATGTCGG 250
TGTCCAAGGA CGTCAAGGTG CCCGTGSACG NTGCAGGCC 289

(2) INFORMATION FOR SEQ ID NO: 28

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 198

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-1089

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 28

TTGNACCANG CCTATCGCAA GCCAATCACC TATGACACGC TGTGGCAGGC 50
TGACACCGAT CCGCTGCCAG TCGTCTTCCC CATTGTGCAA GGTGAACTGA 100
GCAANGCAGA CCGGACAACA GGTATCGATA GCGCCGAATG CCGGCTTGGA 150
CCCGGTGAAT TATCAGAACT TYGCAGTCAC GAACGACGGG GTGATTTT 198

(2) INFORMATION FOR SEQ ID NO: 29

(i) SEQUENCE CHARACTERISTICS:

GCCACCGGTT CATCGCGTGG TGCTGGTCAC CGCCNGGAAN GCCTCAGCGG 50
ATCCCCTGCT GCCACCGCCG CCTATCCCTG CCCCAGTCTC GGCGCCGGCA 100
ACAGTCCCGY CCGTGCAGAA CCTCACGGCT NCTHCCGGGC GGGAGCAGCA 150
ACAGGTTCTC ACCGGYGCCW NGYACCCGCA CCGATCGCGT CGCCGATTCC 200
GGTCGGA 207

(2) INFORMATION FOR SEQ ID NO: 25

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 204

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-1025

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 25

TTNCGCANNC GTTCATCCAG GTCCACTGGT GTCGCANCTC TCNNTGATGC 50
ACCGGTTCCG GATATATGTC NACATCNCCS TCSTCGTCCT GGTGCTGGTA 100
CTNACGAACC TGATCGCGCA TTTCACCACA CCGTGNGCGA GCATCGCCAC 150
CGTCCCGGCC GCCYGCGGTC GGACTGGTGA TCTTGGTKCG GAGTAGAGGC 200
CTGG 204

(2) INFORMATION FOR SEQ ID NO: 26

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 207

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-1035

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 26

ATACCNGTCA TCCNGCACAT NGTCAACCTN GAGTCGGTNC TCACCTACGA 50
GGCACGCCCG AGATGCATCA CTGGTGCTCG RTCAGNCCTT CACGGCTTGG 100
CCGCCTTCCG GTAGGACCGT HGCATGCCCG TCTTCGGCGC CTCGGGTGTT 150

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#3-9

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 31

CAGNCCGCTG NCCCGGAACT GTTCCAGCAG CTACAAGACG TTCGACAACG 50
TNGCGCGTCA ACCTGCANTC GAGCGCAACC TCTCGGTGGC GCTCAACGAG 100
TGTTCGCCGG CTTCAACCCG CTGGACCCGC GAAACCTCGA CGTGTCCCCG 150
CTGCCTTCGC TGGCCAAGCG CGCCGCCGAC ATCCTGCGCC AGGACGTGGG 200
CGGGCAGGTC GACATTTTCG ATGTCAATGT GCCCACCATC CAGTACGACC 250
AGAGC 255

(2) INFORMATION FOR SEQ ID NO: 32

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 164

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#3-12

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 32

AAYNCCNGGC CRTCGACGGT NCCGGTTCNC RCCACCGGTC TATATCCACC 50
CGGGTCNRCA TTMANANTGA NTMNCCGCCG GTGCGGCCGT CGAGCGTGAC 100
CTGGCATCCC CTGAGACGCT GCTGGGTTGC CCCGGGGAGN TCGAMANTCG 150
GGCATCGCAC CATC 164

(2) INFORMATION FOR SEQ ID NO: 33

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 237

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#3-15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 33

(A) LENGTH: 149
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#2-1090
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 29

TCACGANGGT RYNACMGCAA CWCGACCGCC ACGTCASGCC GCCGCGCACG 50
AAGATCACCG TGCCTGCNCG ATGGGTCGTG AACGGAATAG AAYGCAGCGG 100
TGAGGTCAAN YGCGAAGCCG GGAACCAAAT CCGGTGACCG CGTCGGCAT 149

(2) INFORMATION FOR SEQ ID NO: 30
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 210
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#2-1104
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 30

GGACCCGCCA AGCATCAGCC GGTCAACAGC CGCCGCCGGT GGCCAAAGTT 50
CGAGCAGCCG CCGGTATCGT GCTCGGCCCG GCTAGACCAA AAACTTTACG 100
CCAGCGCCCG AAGCCACCCG ACTCCAAGGC CTCGGCCCGG TTGGGTTCGC 150
ACATGGGTGA GTTCTATATG CCCTACCCGG GCACCCGGTT CAACCAGGAA 200
ACCGTCTCGC 210

(2) INFORMATION FOR SEQ ID NO: 31
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 255
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:

GAGAACTCCG GGCCGANTTT TGGACA 26

(2) INFORMATION FOR SEQ ID NO: 36

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 204

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#3-133

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 36

TGTCGGGTNA RNGTTCGCGT CCATGATTGC TCTTGCAACG CTGTTGACGC 50

TTATCAATCA AGTCGTCGGC ACTCCGTATA TTCCCGGTGG CGATTCTCCC 100

GCCGGGACCG ACTGCTCGGA GCTGGCTTCG TGGGTATCGA ATGCGGCGAC 150

GGCCAGGCCG GTTTTCGGAG ATAGGTTCAA CACCGGCAAC GAGGAAGCGC 200

CTTG 204

(2) INFORMATION FOR SEQ ID NO: 37

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 312

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#3-134

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 37

CANNTTAGAC TGTCGTGACA TATCNCNNTN TACNCNTGGN ACGGCCATNA 50

TTGGATAATN CGTGATAANC ACCACAAGAA TNATTCCTAT GNATATTGTC 100

GGTACGTTCG CGNCCATGAT TNGCTCTTGC AACGCTGTTG ACGCTTATCA 150

ATCAAGTCGT CGNCACTCCG TATATTCCCG GTGNCGATTC TCCCGCCGGG 200

ACCGACTGCT CRGAGCTGGC TTCGTGGGTA TCGAATGCGS CGACGSCCAG 250

GCCGGTTTTC GSAGATAGGT TCAACACCGG CAACGAGGAA GCGCCTTGGC 300

GGCTCGGGGC TN 312

(2) INFORMATION FOR SEQ ID NO: 38

ACGGACGGCA ACGGGATGCG ACCCGATCCC ACCGGTCGCC ACGAGGGACG 50
CTACTTCGTC GCCGGGCAGC CGANCCGACC GTCNGTTCNG CGANGGCGAC 100
NGCCGAAGCC GTTGACCCAC NTTGGTCAGC AGCAGCTGGA TSAGTCAGGT 150
GCCGTTGGTG TTTCGCCGTC AGCGGTGTCG GGGTGGGTGC GTTCTGGGCA 200
CCGTCGACTG TGGTGGGCGC TNGCGGGCGN TGGTGGC 237

(2) INFORMATION FOR SEQ ID NO: 34

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 374

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#3-47

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 34

CNGATNGCTC GGNCTNCGGT ACCNAACTCG NAACTCGCGC CCWYGCGNAC 50
GCAGGNCCGC GGTTGGCACC ACCAGCGACA TCAATCANGC AGGWKNCCCG 100
CCACGTTGCA AGACGGCGGC AATCTTCGCC TGTCGCTCAC CGACTTTCCG 150
CCCAACTTCA ACATCTTGCA CATCGACGGC AACAACGCCG AGGTCGCGGC 200
GATGATGAAA GCCACCTTGC CGCGCGCGTT CATCATCGGA CCGGACGGCT 250
CGNACGNACG GTCGACACCA ACTACTTCAC CAGCATCGAG CTGACCAGGA 300
CCGCCCCGCA GGTGGTCACC TACACCATCA ATCCCGAGGC GGTGTGGTCC 350
GACGGGACCC CGATCACCTG GCCG 374

(2) INFORMATION FOR SEQ ID NO: 35

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#3-78 (overlaps with AciI#3-167)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 35

GTGNGCGCGC CNTCGAGCAN GTCTTGGCNG CGANCCCGAB ACAANTGATT 50
CCCGACATCC GGTACACACC GAACCCCNAA NCGATGCGCC NGGCGGCCCG 100
CTGGTAGAAA GGGGAAATCG CCAGTGCTGA CTCGCKTCAT CCGACGCCAG 150
TTGAKCCKTT TKGCGAKCGT CKCCGTAGTG GCAATCGTCG TATTGGGCTG 200
GTACTACCTG CGAATTCCGA GTCTGGTGGG TNGTCGSGCA GTACACCTTG 250
AAGGCCGACT TGCCCGNATC GGGTGGCCTG TATCCGACGG CCAATGTGAC 300
CTACCGCGGT ATCACCATTG GCAAGGTTAC TGCCGTCGAG SCCACCGACC 350
AGGGCNGCAC GANGTGACGA TGAGCATCGC CAGNCAACTA SAAAATCSCC 400
GTCGATGCCT NCGGCGAACG TGCATTCGGN GTCAGCGGTN GGCGAGCAGT 450
ACATCGACCT NGTGTCCACC GGTGCTCCGG GTNAAATACT TCTCCTCCGG 500
ACAGACCATC ACCAANGGCA CCGTTCCCAG TGAGATCGGG CCGGCGCTGG 550
ACAANTCCSA ATCNGCGGGT TGGCCGCATT NGCCCACGGA GAAGATCGGC 600
TTGCTGCTCG ACGAGACNGC GCAAGCGGTG GGTGGGCTGG GACCCGCGNN 650
TTGCAACGGT TGGTCGATTC CACTCAAGCG ATCGTCGGTG ACTTCAAAAC 700
CAACATTGGC GACGTCAACG ACATCATCGA GAACTCCGGG CCGATTTTGG 750
ACAGCCAGGT CAACACGGGT GATCAGATCG ACGCTGGGCG CGCAAATTGA 800
ACAATSTGGC CGCACAGACC GCNGACCAGG GAKCAGAACG TGCGAAGCAT 850
CCT 853

(2) INFORMATION FOR SEQ ID NO: 40

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 209

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#3-204

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 40:

GCGGTTGGCA CCACCAGCGA NAATCAGCAG GNDCCCGCCA CGTTGCAAGA 50
CGGCGGCAAT CTTCGCCTGT CGCTCACCGA CTTTCCGCCC AACTTCAACA 100
TCTTGCACAT CGACGGCAAB AABGCCGAGG TCGCGGCGAT GATGAAAGCC 150
ACCTTGCCGC GCGCGTTCAT CATCGGACCG GACGGCTCGA CGACGGTCGA 200
CACCAACTA 209

(2) INFORMATION FOR SEQ ID NO: 41

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 166

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 676
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#3-166
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 38

AGGCCAATCG NTGATGCGAC TCGAACGGGT TCGGCGCCGA TGACTGTTTC 50
GCGAAGTTCA TCAGCACCCT CGTTGGCGCG AAGGGCACGA CGGTGTACCG 100
GWWRYSAMKA CRCYGCYATG AGTYTCTGCS TGTATTGCGG TGCSGAGCTT 150
GCCGACCCGA CCAGGTGCGG KGCGTGNCGG CSCAKACWAG ATTGGTTCAA 200
CCTGGCNATC GGACCNACGA CGCCGACGGT CGGCGCCGCG ACGACGGCAN 250
ACGGNATNGC GACCCGANTC CNYACCNGGT CGCCACGAGG GACGNCTACT 300
TCGTCGCCNG GCAGCCGACC GANCTCGTTN NNCGCGASGN CGACGCCGAA 350
GCCGTTGACC CACTTGGTCA GCAGCAGCTG GNNATCANGN TCANGGTGCC 400
GTTNNGGTGT TTCGCCGTCA GCGGTGTCGG GGTGGGTGCG TTCTGGGCAC 450
CGTCGACTGT GGTGGGCGCT TGCGGGCGTG GTGGCGTTTC TCGGGCTGGT 500
GGGAGCCGGT GTCGTCGGGA CGCTGTTCCT GAATCGAGAC CGGGAGTCCA 550
TCGACGACAA GTACCTCGCN CCTTGAGGCG GTCCGGACTC ACCGGTGAGT 600
TCAACTCCGA CGCGAACGCC ATCGCCCGCS GCAAGCAGGT GTGCCGCCAG 650
TTGCANASAC GGTGGCGAAC AGCNSA 676

(2) INFORMATION FOR SEQ ID NO: 39
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 853
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#3-167
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 39

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#3-281

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 43

CGGYCCGNNC AAYYYGNCGC GCHNCGGYGY AGAGGTCGNY AAGGTCGCCA 50

AGGTAACGCT GATCGAYGGG NACANGCAAG TATTGGTGNA CTTCACCGTG 100

GHTHGCTHGC TGTYAGC 117

(2) INFORMATION FOR SEQ ID NO: 44

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 385

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: BsaHI#1-21

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 44

GAACCTCCTC GCCCGCGCTT GGCCTAGCAT TAATCGACTG GCACGACAGT 50

TGCCCGACTG GGTACACGGC ATGGACGCAA CGCGAATGAA TGTGAGTTAG 100

CTCACTCATT AGGCACCCCA GGCGTTGACA CTTTATGCTT CCGGCTCGTG 150

TAGTTGTGTG GGAATTGTGG AGCGGATAAC AATTTCGACG ACGAGGAAAC 200

AGCTGTAGAC ATGGATTGAC GAATTTGAAT ACGACTCACT ATAGGAATTC 250

GAGCTCGGTA CCCGGGGATC CTCTAGAGTC CTTCGCCGCG GGTCGCCACC 300

ATCAGGGCCA GTGCGATCGC AAGCGCGGGG TACCGGGCGC CATAGTCTTC 350

AGCATCGGCG TGTTGACCGC AGAGACCGGA CGGGG 385

(2) INFORMATION FOR SEQ ID NO: 45

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 285

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#1-12

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#3-206
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 41

AGATCGTCAG TGAGCAGAAC CCCGCCAAAC CGGCCGCCCG AGGTGTTGTT 50
CSAGGGCTGA AGNCNCTGCT CGCGACGGTC GCTGCTGGCC GTCGTCGGGA 100
TCGGGCTTGG CTCGCGCTGT ACTTCACGCC GGCGATGTCG NCCCGCGAGA 150
TCGTGTATCA TCGGGT 166

(2) INFORMATION FOR SEQ ID NO: 42
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 221
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#3-214
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 42

CCAGNTCCTC NNATATCGAC ACCCTCNACN AAGACCGCTT CGCGAGATCA 50
ACNCTCAGAT ATNCNNACTA TCNCCNNTNC ACGCACACCT CAACATNANA 100
NAATNGAACT ATNGNCTTCG CCTCACCACC AAGGTTCAGG TTANCGGCTG 150
NCGTTTKCTC TKCGCCGGCT CGAACACGCC ATCGTGCGCC GGKACACCCG 200
GATGTTTGAC GACCCGCTGC A 221

(2) INFORMATION FOR SEQ ID NO: 43
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 117
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:

TGGGTGATCG CAGTGAGCAG GTCGACCGCC TATTGGTCAA CGCTAAGACC 200
CTGATCGCCG CGTTNCAACR GASNGCGCCG CGCGGTCGAC GCCCTGCTGG 250
GGAACATCTC CGCTTTCTCG CCCAGGYGCA AAACCTTCAT SAACGACAAN 300
CCGAACCTGA ACCATGTGCT CGAGCNGCGC ATCCTSACSA CCTGTTGGTS 350
GACSGCAAGG AGGATTTGGC TGAAANCCTN ACGATSTTGG GCAGAKTCAG 400
CG 402

(2) INFORMATION FOR SEQ ID NO: 48

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 468

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#1-200

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 48

AGNCCGTGCA CTGGAANCTT CGGCTCAGWT GTCTCCGATG TGGACGGCAA 50
SGCTGATGAT CTCCCGGTTG GAAGTCGANT CGATKASAAA TGGCTTGGCG 100
GCTGGTGGTG TTCGATGCCT GGCACCRACT GGCBACGATC NSCGCCTGGN 150
CGCGATCGGC GCTTAGCTCG GCTGGNNCCC TGTGGTGGGT TTCGACGTGC 200
TCGGTGTTGG TGCTGCTGGT GGTCGAAGGT GTGGCAATCA ACGTTCTGGC 250
TGTTGCGTCG TGATTCGGTA ACCGTCGGTA CCGACGACGA TGCGCCCGGG 300
CTGCGACTGG CCGTTGTCTT CCTGTGCNNG CCGCCGCGAT CTCGGCGGCN 350
GTGGTGACTG GGTACCTGCG CTGGACGACA CCGGACCGCG ACTTCAATCG 400
GGATTCCCGG GAAGTGGTGC ATCTTGCCAC GGGGATGGCC GAGACGGTCG 450
CGTCATTCTC CCCGAGCG 468

(2) INFORMATION FOR SEQ ID NO: 49

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 417

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 45

CCCGCAGCAG TACCCGCAGN CCCACACCCG CTATNCGCAG CCCGAACAGT 50
TCGGTGCACA GCCCACCCNA GCTCGGCGTG CCCGGTCAGT ACGGCCAATA 100
CCAGCAGCCG GGCCAATATG NCCAGCCGGN ACAGTNACGN CCAGCCCGGC 150
CAGTACGCNA CCGCCCGGTC AGTACCCCGG GCAATACGGC CCGTATGNCC 200
AGTCGGGTCA GGGGTCGAAG CGTTCGGTTG CGGTGATCGG CGGCGTGATC 250
GCCGTGATGG CCGTGCTGTT CATCGGCGCG GTTCT 285

(2) INFORMATION FOR SEQ ID NO: 46

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 186

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#1-142

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 46

GCNCGTGNCC GTGCCGCCCG GTTGAACGTG AGCNGCTGNC NATNGCCCCA 50
GCCGAGACGA GAACGTCCCC GAGGAGTATG CAGACTGGGA AGACGCCGAA 100
GACTATGACG ACTATGACGA CTATGAGGCC GCAGACCAGG AGGCCGCACG 150
GTCGGCATCC TGGCGACGGC GGTTGCGGGT NCGGTT 186

(2) INFORMATION FOR SEQ ID NO: 47

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 402

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#1-144

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 47

GTCGCTGAAT GTGTTGTCGG AGACCGTGAT CAGACCTATC CGCACCTGAG 50
CGCCGCCTCC ACGGGTGGCT AAGTTCTCCG ACACCATCGG CAAGCGCGAC 100
GAGCAGACTC ANGCACCTAC TAGCCCAGGC CAACCAGGTG GCCAGCATCC 150

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#2-145

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 51

CGGCCCGGCG GCGCCCTGGT GAAGCTTGGA GAATGGGTGA GCGCAGCTGC 50

CCACCACACG GGACCGGTGC GGACGCGSTG ACGCGCCTGG TGGTCAGCAN 100

CNTGGCCGGT CTGCTGTTGT ATGCCAGCTT CCCGCCGCGC AACTGCTGGT 150

GGCGGCGGTG GTTGGGCTNC GCATTGCTGG CCTGGGTGCT GACCCACCGC 200

GCGACGACAC CGGTGGGTGG GCTGGGCTAC GGCCTGCTAT TCGGCCTGGT 250

GTTCTACGTC TCGTTGTTGC CGTGGATCGG CGAGCTGGTG CNCCGGGCCC 300

TGGTTGGCAC TGNCGACGAC GTGC 324

(2) INFORMATION FOR SEQ ID NO: 52

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 229

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#2-150

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 52

CCAGGCTAGC ACGTATGCTC CGGCTCGTTG TGTGTGGAAT GTGAGCGGAT 50

GACANKNCAC ACAGGADAYA GCTATGACNA TGATTACGCC AAGCTATTTA 100

GGTGABACTA TAGAATAYTC AAGCTATGCA TCCAAYGCGT TGGGAGCTCT 150

YCCATATGGT CGACCTGCAY GCGGCCGCAC TAGTGATTST THGCGCCGGC 200

NYGCWGCGGC NYAYGACCGC YAAYACCAC 229

(2) INFORMATION FOR SEQ ID NO: 53

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 293

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#3-28

(D) OTHER INFORMATION: HinPI#2-23

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 49

GTCCAAGGCC GTAGCCCACC TCCTGGAAGT CGTACCACGT CGACTCGACC 50
AGGACGGCTG CAGTCAGCAC TTCGTCAACC CGCGATCATC AACGTGCACC 100
TACGGCAGTG TGACGCACCC CGGACCATCG CACTGGCCGG GGTTCACACG 150
CCGAACACTG CTGACCGCAC TGGATCTGCT GGTCGCATGC ACCACTTCAA 200
GGTGGTGACG TACCTCAAAA TGGGTTTCCC GTTGTCCACC GAGGAAGTCC 250
CGCTGATTCA TGGGCAATAA CGCTCCCTAT CCGCAGTGTC ACCAGTGGGT 300
GCAAGCGGCG ATGGCCAAGT TGGTCGCTGA CCACCCCGAC TACGTTTTCA 350
CAACCTCGAC TCGACCGTGG AACATCAAAC CCGGCGATGT GATGCCAGCA 400
ACCTATGTCG GGATCTG 417

(2) INFORMATION FOR SEQ ID NO: 50

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 279

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#2-143

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 50

CGGTCGAGCC GATGAACGTC TGCAGTTCAC CGCAACCACG CTCAGCGGTG 50
CTCCCTTCGA TGCGCAAGCC TGCAAGGCAA TGCCGCGGTG TTGTGGTTCT 100
GGACGCCGTG GTGCCCGTTC TGCAACTGTC AGAAGCCCCC AGCCGCAGCC 150
AGGTAGCGGC CGCTAATCCG GCGGTCACCT TCGTCGGAAT CGCCACCCGC 200
GCCGACGTCG GGGCGATGCA GAGCTTTGTC TCGAAGTACA ACCTGAATTT 250
CACCAACCTC AATGACGCCG ATGGTGTGA 279

(2) INFORMATION FOR SEQ ID NO: 51

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 324

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(A) LENGTH: 117
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: HinPI#3-34
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 55

CAGCCACCTC GTTCGCCGCC GACATCGACT ATCAGCCGAC CCGGCCACTG 50
CTGACCTGAT CGCCAACAGC TGGAGGCCCT ACCGGCTGCA GTTCAATTCA 100
CCCGCTGCGG GTCGGCG 117

(2) INFORMATION FOR SEQ ID NO: 56
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 242
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: HinPI#3-41
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 56

AGGTGTCGTG CTTCATGCCT GGCGCCCAAT CCAGTTTCTA CACCGACTGG 50
TATCACCCTT CGCAGACAAA CGGCCAGAAC TACACCTACA AGTGGGAGAC 100
CTTCCTTACC ACACAGATGC CCGCCTGGCT ACAGGCCAAC AAGGCGTGTC 150
CCCCACAGGC AACGCGGCGG TGGGTCTTTC GATCTCGGGC GGTTCCGCGC 200
TGACCCTGGC CGCGTACTAC CCGCAGCAGT TCCCGTACGC CG 242

(2) INFORMATION FOR SEQ ID NO: 57
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 340
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 53

CCACACAACA CAAATCTACG TCGTAATGCA GTCGTAAGTC CATCCGACGT 50
CGATGGCAAG GACAGCACCC GACGGCCAAC GGCATATACA TCGTCGGCTC 100
GCCGGTCACA AGCACATCAT CATGGACTCG TCCACTACGG CGTACCCGTC 150
AACTCGCCCA ACGGATATCG CACCGATGTC GACTGGCCAC CCAGATCTCC 200
TACAGCGGTG TCTTCGTGCA CTCAGCGCCG TGGTCGGTGG GGGCTCAGGG 250
CCACACCAAC ACCAGCCATG GCTGCCTGAA CGTCAGCCCG AGC 293

(2) INFORMATION FOR SEQ ID NO: 54

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 816

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#3-30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 54

CGNCGYCGSC GNGCSCTAYC GGTGCGGGAG GGTACAYCCA AGCANTCCGG 50
GACCGGCCGT CYCGCYGGGA ACGCCGTGCT CCTACAYACC GGCGRCGGGC 100
GCGTTGCCAC GSCCCGACAC CCCACTACCC NGNCGCGGGC GCCACCRTTG 150
GCCCGTTNMG GTGGACCCGA NCTTCCCGGC ACCGCTCGAT GTCCAGCCGT 200
CGCCGCCTAA TCCCGATGGG CCGCMGCCGA CKCCGGGCAT CCTAAGTGCT 250
GGGCGGCCGG GCGAGCCGGN TCCGGNTGTT CCGGCATACC GWTGCCSYTG 300
CCGNCGAACN TGCACGCACC CAACCGCTTG AGCCGTTTCC TGACGGGACG 350
GGAGGTAGCA ACCAATGAGC ACCATCTTCG AYATCCGSAG CCTGCKACTN 400
GYCGAWACTG TCTNGCAAAG GTAGTGGTCG TCGGCGGGTT GGTGGTGGTC 450
TTGGCGGTCG TRGCCGNCTG NCRGCCGGCG CGCRGCTCTA CCGGAAACTG 500
ACTANACTAC CGTGGTCGCR TATTTTCTST GAGGCGCTCG CGCTGTACCC 550
AGGAGASAAA GTCCAGATCA TGGGTGTGCG GGTCGGTTCT ATCGACAAGA 600
TCGAGCCGGC CGGCGACAAG ATGCGAGTCA CGTTGCACTA NCAGCAASAA 650
ATACCAGGTG CCGGCCACGC TACCGNYGNW CGMTCCTCAA CCCCAGCCTG 700
GTGGCCTCGC GCACCATCCA GCTGTCACCN NCGTACACCG GCGGCCCGGT 750
CTTGCAAGAC GGCGCGGTGA TSCCAATCGA GCGCACCCAG RTGCCCGTCG 800
AGTGGGATCA GTTGCG 816

(2) INFORMATION FOR SEQ ID NO: 55

(i) SEQUENCE CHARACTERISTICS:

(ix) FEATURE:
(D) OTHER INFORMATION: HpaII#1-10
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 59.

CCACCANNNA ACRRCACAGC TCCGGCCRRC CGTNCGCAGG CCACCCGCAN 50
CGTAGTGCTC AAATTCTTCC AGGACCTCGG TGGGGYACAT CCGTCCACCT 100
GGTACAAGGC CTTCAACTAC AACCTCGCGA CCTCGCAGCC CATCACCTTC 150
GACACGTTGT TCGTGCCCGG CACCACGCCA CTGGACAGCA TCTACCCCAT 200
CGTTCAGCGC GAGCTGGCAC GTCAGACCGG TTTCGGTGCC G 241

(2) INFORMATION FOR SEQ ID NO: 60
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 243
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: HpaII#1-13
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 60

CCGGCGGATC TGCGTGACGA NTGTATNCCA CGGNACTACC CGCGGTCCTT 50
CCTCNANTNC CGCCGGNCCA GNCGCAGNCT NCNGATGTCC NGCTATAACC 100
TGCGCGATCG CCGCCGGGCT GCCCGACAAC ACGGTGNGCG CCGCCGCTGC 150
TTCCGCCAAT TCTGGGTGNC GGCATNCCGG CAGCGCCCGG CCCAGCACTG 200
AGAGGGGGAC GTTGATGCGG TGGCCGACGG CGTGGCTGCT GGC 243

(2) INFORMATION FOR SEQ ID NO: 61
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2346
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#2-825
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 61

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HpaII#1-3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 57

TGCTGCAGAT AGCCAAGGAT CCAGTCGTGA TTGATATCAC GTCTTTCCAG 50
TGAATTGAAG TTTGGCTATC AAAGGGTGAA CTTSAAAGAC GGCACACTGA 100
CCTATGATGG TGCCGATCCG GAGCGCAAGC GCGCCATGGT TTCCAAGCCA 150
GAGGGCAAGN ACAAGTACGG CGAAGAGCTG GTCGGGCCGG TGCGCGGGCT 200
CAACACCGAG GACCGGACCT ACCTGAATTT CGACAAGGTC GAGACGTTGG 250
GCAGCAGCAC CGAAATTCCG GTGCTGGTGC TGCCGTCCGG CAAGCGTATC 300
GAATTCCAAA TGGCCTCAGC CGATGTGATA CACGCATTCT 340

(2) INFORMATION FOR SEQ ID NO: 58

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 262

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HpaII#1-8

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 58

CNGACTCCAA CNAGTGCGNT CAANCNGNTG TNCCNGACAA GAAGGTTCCT 50
ACATCCGCAA NTCGGTGNAA NGCCACTGTG GATGCCTACG ACGGAACGGT 100
CACGCTGTAC CAACAGGACG NAAAAGGATC CGGTGCTCAA GGCCTGGATG 150
CAGGTCTTCC CCGGCACGGT AAAGCCTAAG AGCGACATTG CGCCGGAGCT 200
TGCCGAGCAN CTGCGGTATC CCGAGGACCT GTTCAAGGTG CAGCGCATGT 250
TGTTGGCCAA AT 262

(2) INFORMATION FOR SEQ ID NO: 59.

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 241

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

CCCAGGCCAT CGTCGCCGGC CGCCACGAGA TCGAGGACGA GCCCGCGGTG 1950
GTGGTGTGGC TGGCGTCCGG CTTGGCCGCC GAGACATTCC AGCTGGACTT 2000
TGTCNGTACC GGCTCGGGTG CCCTGATCAC CGGTTATCGG TTCGACCGNA 2050
CCGCCCGGGA TCTGCATCTG CTGCTGCCGG ACCCGTACAC ATTCCCGTCG 2100
AACCTGCTCA TCGAGCACCC CAACACCGAC CTGCCGGGCA CCGCNGTCGT 2150
GGGCGGCGNT GGTGAGCGGC GGGCGCCGGC GGGGCGACAC CCGGSTGTKC 2200
CGCGATCACG ACGTGCTCAC CTCCGGMGTC GTCGGCGTGC GCCTGCSCGG 2250
GATGCGCGGT GTMCCGGTCG TGTCGCAGGG TTGNCGGCCG ATCGGCTACC 2300
CATACATCGT CACCGGMGCG GACGGCATAC TGRKCACCGA GCTCGG 2346

(2) INFORMATION FOR SEQ ID NO: 62

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 841

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#435

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 62

CGTTACCCGC TTTACACCAC CGCCAAGGCC AACCTGACCG CGCTCAGCAC 50
CGGGCTGTCC AGCTGTGCGA TGGCCGACGA CGTGCTGGNC NAGSCCNANS 100
CCAATGNCGG MMTGCTGCAA NCGGNTNCNG GCCANGCGTT CGGACCGGAC 150
GGACGCTGGN CGGTATCAGT CCNGTCGGCT TCAAANCCGA NGGCGTGGGC 200
GAGGACCTCA AGTCCGRRCC CGGTGGTCTC NAAACCCSGG CTNGTCAACT 250
CCGATNCGTC GCCCAACAAN CCCAACNGCC NGCCATCANC GACTCCKCNG 300
GCACCGCCNG AGGGAAGGGY CCGGNTCGGG ATTCAACGGG TTGGCRWCGC 350
GGCGCTGCCG TTCNGRATTG GAYCCGGCAN CGTACCCCGG TGATGGGCAG 400
CTNACGGGGA NGAACAACCY GSCCSSSACG GCCACCTCGG CCTGGTACCA 450
GTTACCGCCC CGCAGCCCGG ACCGGCCNGC TGGTGGTGGT TTCCNGCGGC 500
CGGCGCCATC TGGTCCTACA AGGAGGACGG CGATDTCATC TACGGCCANG 550
TCCCNTGAAA CTGCAGTGGG NCGTCACCGG CCCGGACGGC CGCANTCCAG 600
CCACTGGGGC AGGTATTTCC GANTCGACAN TCGGACCNGC AACNCCNGCG 650
TGGCGCAATC TGCGGTNTNT CCGCTGGCCT GGGCGCCGCC GGNANGCNCG 700
ACGTGGCGCG CATTGTCGCC TATGACCCGA ACCTGAGCCC TGAGCAATGG 750
TTCGCCTTCA CCCCGCCCCG GGTTCCGGTG CTGGAATCTC TGCAGCGGTT 800
GAKCGGGTCA GCGACACCGG TGTTGATGGA CATCGCGACC G 841

GCGCTGNCAT TCGNACTTCG GACNGCGTTN GCGGTGGTGC TGATCATGAA 50
NCTACGACGG CGCCACCGGC AGCTTCCCGT CATGGGTGCT CTATCCCTGT 100
GCGCTGGCCA TGATGGTGTT CTCGAATKCG TTCAGCGTNC TGCGCAGCGC 150
AGTGANACCG AGGGTGATGC CGCCAACCAT CGACTTGGTC CGGGTCAACT 200
CACGGCTGAC CGTGTTCGGC CTGCTCGGCG GCACCATCGC TGGTGGCGCG 250
ATTGCGGCCG GAGTCGAATT CGTCTGCACC CACCTGTTCC AGCTGCCGGG 300
CGCGTTGTTC GTCGTCGTCG CGATCACCAT CNNTNNNGCT TCGCTGTCGA 350
TNCNCATTCC GCGCTGGGTC GAGGTGACCA GCGGTGAGGT CCCGGCCACA 400
TTGAGCTACC ACCGGGATAG GGNCAGACTA CGGCGACNGC TGGCCGGAGG 450
AAGTCAAGAA CCTCGGCGGA ACACTCCGAC AACCGTTGGG CCGCAACATC 500
ATTACCTCCC TGTGGGGTAA CTGCACCATC AAGGTGATGG TCGGCTTTCT 550
GTTCTTGTAT CCGGCGTTTG TCGCCAAGGC GCACGAAGCC AACGGGTGGG 600
TGCAATTGGG CATGCTGGGC CTGATCGGCG CGGCGGCCGC GGTCGGCAAC 650
TTCGCCGGCA ATTTCACCAG CGCACGCCTG CAGCTAGGCA GGCCAGCTGT 700
GCKGGTNGTG CGCTGCACCG TGCTAGTTAC CGTGTTAGCC ATCGCGGCCG 750
CGGTGGCCGG CAGCCTGGCA GCGACAGCNA TTGCCACCCT GATCACGGCA 800
GGGTCCAGTG CCATTGCTAA AGCCTCGCTG GACGCCTCGT TGCAGCACGA 850
CCTGCCCGAG GAGTCGCGGG CATCGGGGTT TGGGCGTTCC GAGTCGACTC 900
TTCAGCTGGC CTGGGTGCTG GGCGGCGCGG TGGGCGTGTT GGTGTACACC 950
GAGCTGTGGG TGGGCTTCAC TGCGGTGAGC GCGCTGCTGA TCCTGGGTCT 1000
GGCTCAGACC ATCGTCAGCT TCCGCGGCGA TTCGCTGATC CCTGGCCTGG 1050
GCGGTAATCG GCCCGTGATG GCCGAGCAAG AAACCACCCG TCGTGGTGCG 1100
GCGGTGGCGC CGNAGTGAAG CGCGGTGTCG CAACGCTGCC GGTGATCCTG 1150
GTGATTCTGC TCTCGGTGGC GGCCGGGGCC GGTGCATGGC TGCTAGTACG 1200
CGGACACGGT CCGCAGCAAC CCGAGATCAG CGCTTACTCG CACGGGCACC 1250
TGACCCGCGT GGGGCCCTAT TTGTACTGCA ACGTGGTCGA CCTCGACGAC 1300
TGTCAGACCC CGCANGCGCA GGGCGAATTG CCGGTAAGCG AACGCTATCC 1350
CGTGCAGCTC TCGGTACCCG AAGTCATTTC CCGGGCGCCG TGGCGTTTGC 1400
TGCAGGTATA CCAGGACCCC GCCAACACCA CCAGCACCTT GTTTCGGCCG 1450
GACACCCGGT TGGCGGTCAC CATCCCCACT GTCGACCCGC AGCGCGGGCG 1500
GCTGACCGGG ATTGTCGTGC AGTTGCTGAC GTTGGTGGTC GACCACTCGG 1550
GTGAACTACG CGACGTNCGC ACGCGGAATG GTCGGTGCGC CTTATCTTTT 1600
GACGAGGCCG CGGCTCGACG NCNCCTTAAG CGCGGTCGGC GCCAACGGTC 1650
CGAAGAGCCG CCGACACCCG GGGCACATCG GCGCATCATG GAACTGTGCG 1700
GATCGGAGTC GGGGTTTGCA CCACGCCCGA CGCGCGGCAG GCCGCGGTGG 1750
AGGCTGCGGG CCAGGCGCGC GACGAGCTGG CGGGTGAGGC GCCGTCGCTG 1800
GCGGTGTTGC TTGGATCGCG TGCACACACC GACCGGGCTG CCGACGTCCT 1850
GAGCGCGGTG CTGCAGATGA TCGACCCGCC CGCGCTTGTC GGTTGCATCG 1900

GNATNCGACG GCAATAAACA CACGCTCTGG CCACGTTTCT TGGCGGGGAA 250
AGGGGTGATG CTATCGGAGC CAATGGTATC GCGACAACAC TTGCAGATGC 300
CGCCAAGGCC GATCACGCTA ATGACGGATT CGGGGCCACA AACGTTCCCC 350
GTTCTGGCGG TTTTCTCTGA CTACACCTCA GATCAAGGTG TGATTTTGAT 400
GGATCGCGCC AGTTATCGGG CCCATTGGCA GGATGATGAC GTGACGACCA 450
TGTTTCTTTT TTTGGCNATN CGGGTGCGAA TAGCG 485

(2) INFORMATION FOR SEQ ID NO: 65

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 469

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#1-264A

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 65

GGCGAGGTCA GTGAAGCCGA GGAAGCGGAA AGGAGCGCCC AATACGGAAC 50
CGCCTCTCCC CGCGCGTTGG CCGATTCATT AAATGCAGCT GGCACGACAG 100
GTTTCCCGAC TGGAAMGCGG GCAGTGAGCG CAASGCAATT AATGTGAGTT 150
AGCTCACTCA TTAGGCACCC CAGGCTTTAC ACTTTATGCT TCCGGCTCGT 200
ATGTTGTGTG GAATTGTGAG CGGATAACAA TTTCACACAG GAAACAGCTA 250
TGACATGATT ACGAATTTAA TACGACTCAC TATAGGGAAT TCGAGCTCGG 300
TACCCGGGGA TCCTCTAGAG TCGCTTCGGT TGGCGGCGAC CAGCAGTGGA 350
TCCACGGTGG CCGCCCGCGC GGCDTCATAC ACCGCCGCGG CCTCCTTGGC 400
CTGTGCGGCC SGCTTAGCGC GCGTGTTGCT GCCGTGCTTA GCCANCTGGC 450
ATAGGGGGCT GCCGCGCGC 469

(2) INFORMATION FOR SEQ ID NO: 66

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 290

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(2) INFORMATION FOR SEQ ID NO: 63

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 471

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#1-2/23/9

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 63

GCCAGCCGTG ATCGGCTGAC CGGCAGTGAT CACCAACCTC AACGTGGTGC 50

TGGGCCTCGC TGGCGCTCAC ACGATCGGTT GGACCAGCCG GTGACGTCGC 100

TATCAGCGTT GATTCACCGG CTCGCGCAAC GCAAGACCGA CATCTCCAAC 150

GCCGTGGCCT ACACCAACGC GCCGCCGGCT CGGTCGCCGA TCTCTGTCGC 200

AGGCTCGCGC CGTTGGCGAA GGTGGTTCGC GAGACCGATC GGGTGGCCGG 250

CATCGCGGCC GCCGACCACG ACTACCTCGA CAATCTGCTC AACACGCTGC 300

CGGACAAATA CCAGGCGCTG GTCCGCCAGG GTATGTACGG CGACTTCTTC 350

GCCTTCTACC TGTGCGACGT CGTGCTCAAG GTCAACGGCA AGGGCGGCCA 400

GCCGGTGTAC ATCAAGCTGG CCGGTCAGGA CATGCGGCGG TGCGCGCCGA 450

AATGAAATCC TTCGCCGAAC G 471

(2) INFORMATION FOR SEQ ID NO: 64

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 485

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#1-229/264

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 64

KGTCTCGCGN CCTTNACATC CGGTCGCCNN RCGGTNATCT GCCTGTGGAT 50

GCCGTCCGGA NGTATNANCN AATGGCCANG AGTNCGTGAC NGCAGNTATG 100

GNCKCGGNTA TAGTTCCGTT TTGCCCNGGA CTNGGNGCGT GAGGTGGAAC 150

TAATGGCGGT GTCGGGTGAT ATTTCCGACG GCAAGNCGAC CATATAGGTG 200

CACTGCGCCG AGATGTGTGG CACGTATCAC TCGATGATGA ACTTCGAGGT 950
CCGCGTCGTG ACCCCCAACG ATTTCAAGGC CTACCTGCAG CAACGCATCG 1000
ACGGGAAKAC AAACGCCGAG GCCCTGCGGG CGATCAACCA GCCGCCCCTT 1050
GCGGTGACCA CCCACCCGTT TGATACTCGC CGCGGTGAAT TGGCCCCGCA 1100
GCCCGTAGGT TAGGACGCTC ATGCATATCG AAGCCCGACT GTTTGAGTTT 1150
GTCGCCGCGT TCTTCGTGGT GACGGCGGTG CTGTACGGCG TGTTGACCTC 1200
GATGTTCGCC ACCGGTGGTG TCGAGTGGGC TGGCACCACT GCGCTGGCGC 1250
TTACCGGCGG CATGGCGTTG ATCGTCGCCA CCTTCTTCCG GTTTGTGGCC 1300
GCGGAT 1306

(2) INFORMATION FOR SEQ ID NO: 68

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 759

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-823

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 68

GGTGCCTGCC ATCGGTTCGC TGNGCCACNG CTGNCNNATC TTTGGTSTGT 50
TAGAGGTNWW CCGCGCGGAT RGCNCANTCC TGTTGGNGGG GGTTRTCGCC 100
ACGATTGCCG CCCGCGCTGA ACCCGACGAC GCCGATGCCC TGCCCACCAC 150
GGATCGGCTG NNMMCANCCG AGCGAACCGT GCAGNATGCN TNTKGTTGAC 200
GAGCCTGCTG GCGCCTTCGC NGGCNCTCGG CGACCATCGG TGCCATCGGA 250
ACCGCCGTNC GCAACCCACG GCATCCACAN GSTCCANGCA TGGCGGTATC 300
GCGNTTGGCC GNCGTCACCG GTGCGCTGCT GCTGCTAYGA GCACGTTCAG 350
CAGACACCAG AAGGTCACTG NTGTTTGCCA TCTGTNGGAA TCACCACCGT 400
TGCAACGGMA NTTGTACCGT CGCCGCGGAT CGGGCTCTGG AACACGGGCC 450
GTGGATTGSC GCGCTGACCG CCATGCTGGT CCNGCCGTGG CAANTGKKTT 500
TGGGCTTCGT NGCTCNCCGC GTTGTCGCTC TCGCCCGTCA CGTACCGCAC 550
CATCGAATTG CTGGAGTGTC TGGCGCTGAT CGCAATGGTT CCATTGACCG 600
CTNTGGSTAT NNNNNCGCCT ANCAGSSSCS TTCGCCACCT CGACCTGACA 650
TGGACATGAC CACNGTCCCG TNACCCTGCG CCTGNCTNGG TGGTMTCAGC 700
GNCNNNTCGY SACGCTGTCT GGSWTGGSRM RCGCNCGGTT GCGCCACGCG 750
GTTTCGCCG 759

(2) INFORMATION FOR SEQ ID NO: 69

(D) OTHER INFORMATION: AciI#1-264C

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 66

CNGGTTCGAC TGATCTAGCT GGGGCCAGAC CGGCACGAGG CGACAGTTAC 50
CAGTACCTGA CAGACAGGCC GATCGAGCCA AACCGTAGTG AGGACGCAGG 100
AGGAACAGGC AGATGCATCT AATGATACCC GCGGAGTATA TCTCCAACGT 150
GATATATGAA GGTCCGCGTG CTGACTCATT GTATGCCGCC GACCAGCGAT 200
TGCGACAATT AGCTGACTCA GTTAGAACGA CTGCCGAGTC GCTCAACACC 250
ACGCTCGACG AGCTGCACGA GAACTGGAAA GGTAGTTTCA 290

(2) INFORMATION FOR SEQ ID NO: 67

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1306

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#2-92

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 67

GTGATACAGG AGGCGCCAAC AGTGACACCT CGCGGGCCAG GTCGTTTGCA 50
ACGCTTGTCG CAGTGCAGGC CTCAGCGCGG CTCCGGAGGG CCTGCCCGTG 100
GTCTTCGACA GCTGGCGCTC GCAGCAATGC TGGGGGCATT GGCCGTCACC 150
GTCAGTGGAT GCAGCTGGTC GGAAGCCCTG GGCATCGGTT GGCCGGAGGG 200
CATTACCCCG GAGGCACACC TCAATCGAGA ACTGTGGATC GGGGCGGTGA 250
TCGCCTCCCT GGCGGTTGGG GTAATCGTGT GGGGTCTCAT CTTCTGGTCC 300
GCGGTATTTC ACCGGAAGAA GAACACCGAC ACTGAGTTGC CCCGCCAGTT 350
CGGCTACAAC ATGCCGCTAG AGCTGGTTCT CACCGTCATA CCGTTCCTCA 400
TCATCTCGGT GCTGTTTTAT TTCACCGTCG TGGTGCAGGA GAAGATGCTG 450
CAGATAGCCA AGGATCCCGA GGTCGTGATT GATATCACGT CTTTCCAGTG 500
GAATTGGAAG TTTGGCTATC AAAGGGTGAA CTTCAAAGAC GGCACACTGA 550
CCTATGATGG TGCCGATCCG GAGCGCAAGC GCGCCATGGT TTCCAAGCCA 600
GAGGGCAAGG ACAAGTACGG CGAAGAGCTG GTCGGGCCGG TGCGCGGGCT 650
CAACACCGAG GACCGGACCT ACCTGAATTT CGACAAGGTC GAGACGTTGG 700
GCACCAGCAC CGAAATTCCG GTGCTGGTGC TGCCGTCCGG CAAGCGTATC 750
GAATTCCAAA TGGCCTCAGC CGATGTGATA CACGCATTCT GGGTGCCGGA 800
GTTCTTGTTC AAGCGTGACG TGATGCCTAA CCCGGTGGCA AACAACTCGG 850
TCAACGTCTT CCAGATCGAA GAAATCACCA AGACCGGAGC ATTCGTGGGC 900

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: HinPI#1-3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 70

AGATCNAYAC YANCANCANT GCNGTCATCG AGNTGCTGCA GGNCANGGTG 50
GTCCGTTGGC GAACGTGCTN KGCCNAYACC GGTGCCTTCT CGGCGCNCTN 100
GGYGCAYNGC GACCAGCTGA TCGGCGNAKG TAATCACCAA CCTCAANNKC 150
GGTGCTNGCK ACCKTCGAYK GCAAAGAGYG YGCAATTTGT CGGCCAGTGT 200
CGACCAGCTG CAGCAGCTGG TCAGCGGCCT GGCCAAGAAC CGGGATNCCG 250
ANTSGNGGGC GCCATTTCGC CGCTGGNGTC GACGACGACG GATCTTWCGG 300
AACTGTTGCG GAATTSGCGC CGGCCGCTGC AAGGCAKCCT GGAAAACGCC 350
CGGCCGCTGG CTACCGAGCT GGACAACCGA AAGGCCNANG GTCAASAACG 400
RRATCGAGCA NGCTCGGCGA GGACNATNCC TGCGCCTGTC CGCGCTGGGC 450
AGTTACGGAG CANTTCGTTC AACATCTAST TSTGCTCGGT GACGATSAAG 500
ATCAACGGAC CGGCCGGCAG CGACANTCCN TGCTGCCGAT CGGCGGCCAG 550
CCGGANTCCC AGCAAGGGGA GGTGCGCCTT TGCNTAAATA GGAAGCCAAG 600
TANGCAAASA CGAASGCSAC CCGTCCGCAC CGGNCATCTT CGGCCTGGTG 650
CNTGGTGATC NTGNCGTCGT CCTGATSGNC ATTCGGCTAC AGCGGGTTGC 700
CTKTCTGGCC ACAKKKCAAA ACCTACGACG CGTATTTCAC CGACGCCGGT 750
GGGATCACCC CCGGTAACTC GGTTTATGTS TCGGGCCTCA AGGTGGGCG 799

(2) INFORMATION FOR SEQ ID NO: 71

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 713

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-827 translation strand

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 71

CTAYCSGCAA NGCTKNGCAG ACGCTCGGCT GCACNGCAGA ANTCGCGGTG 50
CACCCACGAT TGCCAGTAGC GCGGGCCCAC TCGTGCCTAC TACACTTCGT 100
CGTAGCCAAA TCANTCGGCC CCGTAGTATC TCCGGAGATG ACAGATGAAT 150
GTCGTCGACA TTTCNGNCGG TGGCAGTTCG GTATCACCAC CGTSTATCAC 200

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1041
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: HinPI#1-31
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 69

GKTCNCGGTG ATGTCGACNG TCGGCACGRM GNCGAAACCT CANCGGTCGA 50
CAGTGTCTGC CCGAGGCCGC AGCCGACGTG CCCCNGGAGA CCGCGCGCCA 100
ANCACGGTGC CGTACATGTA GCCCGCACGG CGCATCATCG CCGAGCCGGC 150
GTAGATGTTT TCCTGCACGG CGTNCSCGGT GAACCCTCCG GCGCCAGCAC 200
CGSCACCWNT TCCCGCGTCC ACGTCGGCCT GGGTGGTGAC GCCGAGCACC 250
CCACCGAAAT GATCGACATG GCTGTGGGTG TAGATGACCG SCGACCACGG 300
GGCGGTCGGC TCCGCGGTGG GCGCGANTAC AAGTCCAGCG CGGCGGCGGC 350
CACCTCGGTG GACANCCAAN CGGGYNYGAT GACGARWCWG CCCAGTGTCA 400
CCNCWMMACG AAGNCTGATA TTGGAGATAT CGAATCCGCG GACCTGATAG 450
ATGCCCGGCA CCACCTGGTA GAGGCCCTGT TTCGCGGTCA GCTGGGATTG 500
CCGCCACAGG CTGGGATGCA CCGATGTCGG CGCGGCACCG TCGAGNAACG 550
AGTACGCGTC GTTGTCCCAC ACCNACGCGA CCATCGGCAG CCTTGATCAC 600
ACACGGGGAC AGCGCGGCAA TGAATCCGCG ATCGGCGTCG TCGAAATCCG 650
TTGTGTCATN GCAACGGTNA ACGAGTGTTC ACCGTGTGCC GCCTGGNATG 700
ACGGCAGTNG GGAGGTTTGT GTTCCATCGG CACTACATTG CCACTACTAC 750
GGTGCACGCC GGTAGATGCC GTTGGCGAAC CACGCTACCG ACCAGAAAGA 800
GAGAATTTTC CGCCGCACCT AGACCTCGGG CCCTCTAACG CGCATACTGC 850
CGAAGCGGTC CTCAATGCCG ATGGACCGCT ACGACAGGCA AAGGAGCACA 900
GGGTGAAGCG TGGACTGACG GNTCGCGGTA GCCGGAGCCG CCATTCTGGT 950
CGCAGGTCTT TCCGGATGTT CAAGCAACAA GTCGACTACA GGAAGCGGTG 1000
AGACCACGNA CCGCGNGCAG GCACGACNGC AAGCCCCGGC G 1041

(2) INFORMATION FOR SEQ ID NO: 70
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 799
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ix) FEATURE:
(D) OTHER INFORMATION: AciI#2-874
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 73
GTGATGCCTT CCAGCATTGG ATTGGTCGTC GGTTCGATGC TGTGGCGACA 50
GATAAACCGC CTGTTCGGGG TGCGTGGCCT CTGCTGGGCA GCGCACTGCT 100
CAACGCCGCT CTGCGCTGCT GTGCATGGTG GCCGAGTCGT GTGGGCAGTG 150
GGTTCACGCC TGGGCGTACT TCACGGCGTT CCTGCTGGCT ACGGTGGCCG 200
CTCAAACGGT GGTCGCCGCA TCGATATCGT GGATCAGCGT CCTCGCGCCC 250
GA 252

(2) INFORMATION FOR SEQ ID NO: 74
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 160
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: AciI#2-1018 .
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 74
GGCGCCGCCG TCGTGCTGGC CGCCCGGCCC GGTGGGGGTG CCGGCCAGCG 50
TGGTTCCGCC AGTGGCCGCG CCGAACGTAT TGGCCGGCGT CCTCGAGCAC 100
GACAACGACG GGTCGGGGGC GGCGGTGCTG GCCGCGCTGG CCAAGCTGCC 150
ACCCGGTGGT 160

(2) INFORMATION FOR SEQ ID NO: 75
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 393
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis
(ix) FEATURE:
(D) OTHER INFORMATION: **HinPI#1-27
(xi) SEQUENCE DESCRIPTION: SEQ ID NO 75

TTNCAWYTTC GTNACSYGYT GACCWWCGGC CTGGCNCNCC TKSTKANYRC 250
GGNTCNAYGC AAACTGCTGT GGTCGTCACC GATAANCCCG CCTGGTATCG 300
CCTCACCNAA ATTCTTCGGC AAATTGTTCC TGNATCNAAC NTTTGCCATC 350
GGCGTGGCGA CCGGAATCGT GCAGGNAATK TCAGTTCGGC ATGAACTGGA 400
GCGAGTACTC CCGATTCGTC GGCGATGTCT TCGGCGCCCC GCTGGCCATG 450
GAGNSCTGGC GGCCTTNCTT CTTCGAATCC ACCTTCATCG GGTTGTGGAT 500
CTTCGGCTGG AACAGGCTGC CCCGGCTGGT GCANTCTNGG CCTGCATCTG 550
GNATCGTCGC AATNCGCNGG TNCAACGTGT CCGCGTTCTT CATCATCGCN 600
GGCAAACTCC TTCATGCAGC ATCCGGTCGG CGCGCACTAC AACCCGACCA 650
CCGGGCGTGC CGAGTTGAGC AGCATCGNTC NGTGNCNTGC TGACCAACAA 700
CACCGCACAG GCG 713

(2) INFORMATION FOR SEQ ID NO: 72

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 274

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: AciI#2-834 translation strand

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 72

CCGCAGCACC GAGGCAAGCA TCGCACCCGT CGATTCCCGC CATCCCGGCG 50
ACATGATGGT CATGTCCGAC ACCGACGCCC GCACCTCGCT TCCCGAGTTG 100
ACCGCGCTGC GCGTGGACGC CGCAACGGAT GCGTCGGTTC ATTCGATCCC 150
GGCTCGAAAT TGGCCATGGC GAACGCATCT TGCTGTGATG GTTCGGGCAG 200
TAGATCTCCA CTGCCGCACT GATAAACTCG GGTCATGGTC GTCGTGAGGC 250
GGACAGGGTA GAGGCGCATG ACCG 274

(2) INFORMATION FOR SEQ ID NO: 73

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 252

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

I claim:

1. An isolated *Mycobacterium tuberculosis* nucleic acid sequence including a sequence selected from the group consisting of Seq. I.D. Nos. 1 - 76.

2. A purified immunostimulatory peptide encoded by a sequence according to claim 1.

3. An antibody that specifically binds to a peptide according to claim 2.

4. A vaccine preparation comprising at least one immunostimulatory peptide according to claim 2 and a pharmaceutically acceptable excipient.

5. A purified immunostimulatory peptide encoded by a nucleotide sequence selected from the group consisting of Seq. I.D. Nos. 1 - 76.

6. A vaccine preparation comprising at least one peptide according to claim 5 and a pharmaceutically acceptable excipient.

7. A purified immunostimulatory *Mycobacterium tuberculosis* peptide, the peptide including at least 5 contiguous amino acids encoded by a nucleic acid sequence selected from the group consisting of Seq. I.D. Nos. 1 - 76.

8. A vaccine preparation comprising at least one peptide according to claim 7 and a pharmaceutically acceptable excipient.

9. A peptide according to claim 7 wherein the peptide includes at least 10 contiguous amino acids encoded by a nucleic acid sequence selected from the group consisting of Seq. I.D. Nos. 1 - 76.

10. A vaccine preparation comprising at least one peptide according to claim 9 and a pharmaceutically acceptable excipient.

11. A method of making a vaccine comprising:

providing at least one purified peptide encoded by a nucleotide sequence selected from the group consisting of Seq. ID. Nos 1 - 76;

combining the peptide with a pharmaceutically acceptable excipient.

12. An isolated nucleic acid molecule having a nucleotide sequence selected from the group consisting of:

(a) Seq. ID Nos. 1 - 76;

(b) nucleotide sequences complementary to a sequence defined in (a); and

(c) nucleic acid molecules of at least 15 nucleotides in length which hybridize under conditions of at least 75% stringency to a sequence defined in (a) or (b).

13. A recombinant DNA vector including a nucleic acid molecule according to claim 12.

14. A transformed cell containing a vector according to claim 13.

ATCAGCCGCG GGTCGACGCC GCCGATGACC TCGACGTCGT CGTCGTCGCT 50
GCCGGTACTC AATCCAATCA CCATCCTCTT ACGCACCTTC TAGGAGTGTG 100
TTGCTGCGGC AGTGCCGGCC ATTCGTAGAT TCGGGCCTCG CCGTTGTCGT 150
AGATCTTCGC CCACGACCTC GATGTCTCTA ACGACACTAG TCCGTCCGGC 200
ACGCAAACCC CGCACCGTCG GAGTGCTGGT CAGGTATAGA CGGTACAGGA 250
GGACTTGGTA GGCCTCGAGT ACCGAGGTAC GTCTCCCGTT GCGGCATAGG 300
CCAGAAGATG AACCGGTGTA GACCGGGCCT GTTGCGAGGG TCGTAGTCGT 350
AGGTCCCAGA GGTGTCGGAC GCCCAGGTTA ATACACAGCG TGC 393

(2) INFORMATION FOR SEQ ID NO: 76

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 248

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(D) OTHER INFORMATION: #2-147

(xi) SEQUENCE DESCRIPTION: SEQ ID NO 76

GCAGACCTCT GGCCGCTGGT GGTGCTGGGT ACCTGCGCTG GCGACACCGG 50
ACCGCAGACC GTCAATCGGG ACTCCCGGGA ACGTGGTGCC ATCTTGCCAC 100
GGGGATGGCC GACGCGGCTC GTCATTCTCC CCGAGCGCAC CGGCCGCCGC 150
TGTTGACCGG GCCGCGGCGA CTGATGGTGC CCGCACACGC GGGCGGGTTC 200
AAGGAGCAAT ACGCCAAGTC CAGCGCCGCT CTCGCACGGC GCGGTGTT 248

15. A nucleic acid probe comprising a nucleic acid molecule according to claim 12 and a diagnostic label.

16. A method of isolating a *Mycobacterium tuberculosis* gene which gene encodes an immunostimulatory peptide, the method comprising the steps of:

providing nucleic acids of *Mycobacterium tuberculosis*;

contacting said nucleic acids with a probe or primer, the probe or primer comprising at least 15 contiguous nucleotides of a polynucleotide having a nucleotide sequence selected from the group consisting of Seq. ID Nos. 1 - 76 and sequences complementary thereto; and

isolating the *Mycobacterium tuberculosis* gene.

17. An isolated *Mycobacterium tuberculosis* gene produced by the method of claim 16.

18. An isolated *Mycobacterium tuberculosis* nucleic acid molecule, said molecule encoding an immunostimulatory peptide and hybridizing under conditions of at least 75% stringency to a nucleic acid probe comprising at least 20 contiguous bases of a sequence selected from Seq. ID Nos. 1 - 76.

19. A purified immunostimulatory peptide encoded by the nucleic acid molecule of claim 18.

20. An immunostimulatory preparation comprising:
a purified peptide according to claim 19; and
a pharmaceutically acceptable excipient.

21. An improved tuberculin skin test, the improvement comprising the use of one or more immunostimulatory peptides according to claim 19.

22. A vaccine preparation comprising an immunostimulatory membrane peptide isolated from *Mycobacterium tuberculosis* and a suitable excipient.

23. A method of detecting the presence of *Mycobacterium tuberculosis* DNA in a sample comprising contacting the sample with a nucleic acid probe according to claim 15 and detecting hybridization products that include the nucleic acid probe.

24. A method of detecting the presence of *Mycobacterium tuberculosis* DNA in a sample comprising:

selecting two or more nucleic acid primer molecules from the nucleic acid molecules defined in claim 12, said molecules suitable for amplification of a *Mycobacterium tuberculosis* target sequence;

incubating the sample under conditions suitable to amplify the target sequence; and

detecting an amplified product.

25. A method of detecting the presence of a *Mycobacterium tuberculosis* peptide in a sample comprising contacting the sample with an antibody according to claim 3 and detecting the presence of an antibody-peptide complex.

26. A method of detecting the presence of an anti-*Mycobacterium tuberculosis* antibody in a sample comprising contacting the sample with a peptide according to claim 2 and detecting the presence of an antibody-peptide complex.

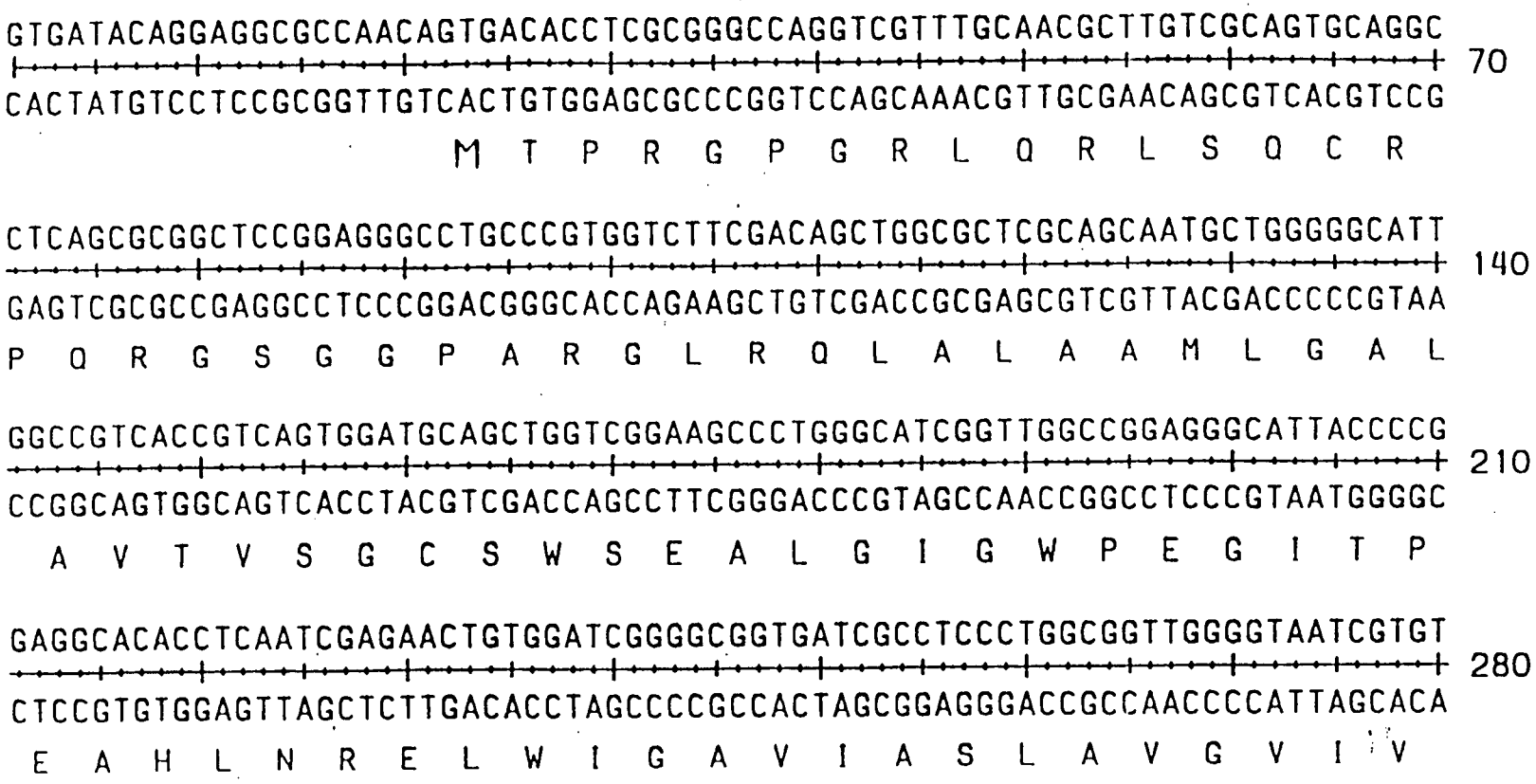

FIG. 1
(Page 1 of 4)

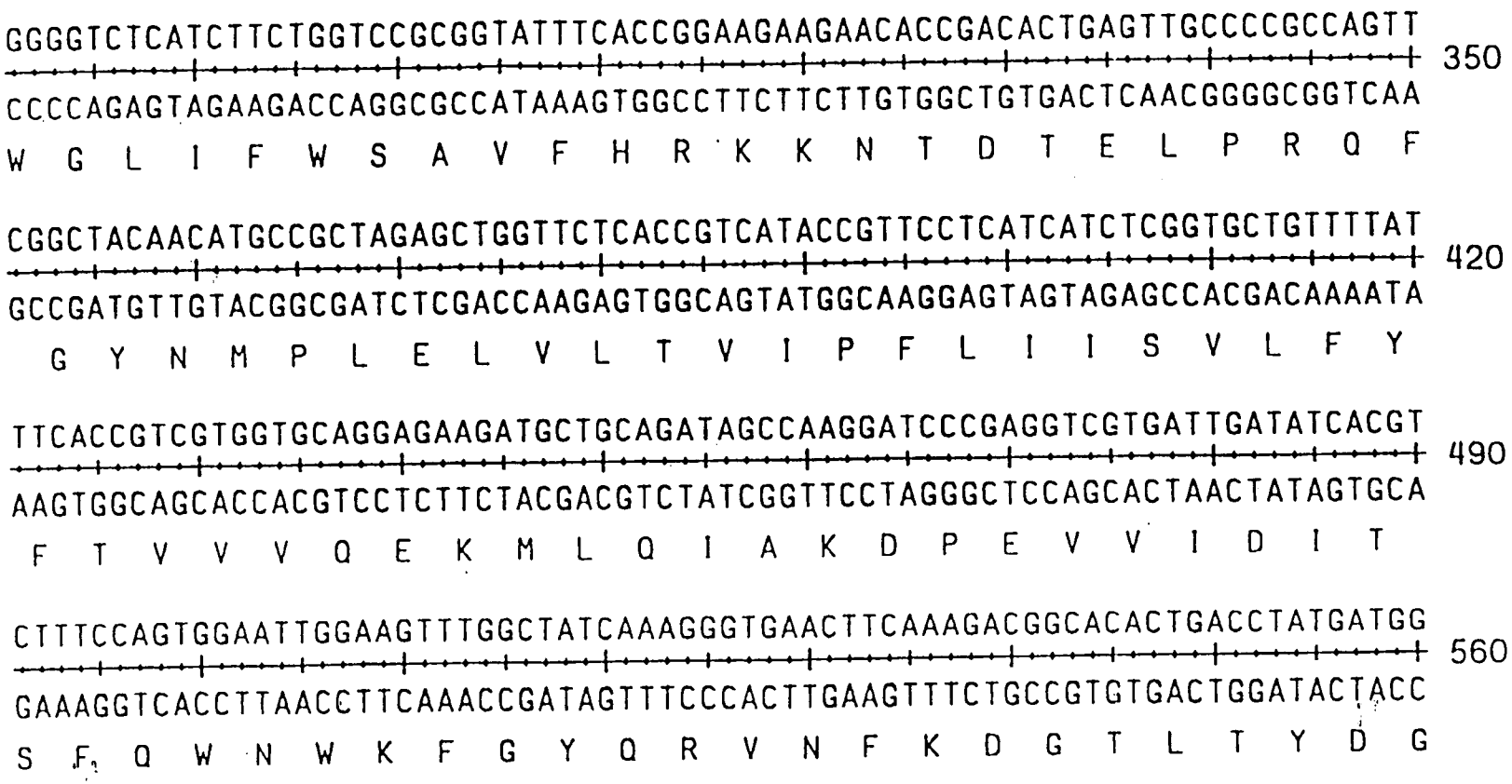

FIG. 1

(Page 2 of 4)

FIG. 1

(Page 3 of 4)

```
AACAACTCGGTCAACGTCTTCCAGATCGAAGAAATCACCAAGACCGGAGCATTCGTGGGCCACTGCGCCG  910
TTGTTGAGCCAGTTGCAGAAGGTCTAGCTTCTTTAGTGGTTCTGGCCTCGTAAGCACCCGGTGACGCGGC
  N  N  S  V  N  V  F  Q  I  E  E  E  I  T  K  T  G  A  F  V  G  H  C  A

AGATGTGTGGCACGTATCACTCGATGAACTTCGAGGTCCGCGTCGTGACCCCCAACGATTTCAAGGC  980
TCTACACACCGTGCATAGTGAGCTACTACTTGAAGCTCCAGGCGCAGCACTGGGGGTTGCTAAAGTTCCG
  E  M  C  G  T  Y  Y  H  S  M  M  N  F  E  V  R  V  V  T  P  N  D  F  K  A

CTACCTGCAGCAACGCATCGACGGGAATACAAACGCCGAGGCCCTGCGGGGCGATCAACCAGCCGCCCCTT  1050
GATGGACGTCGTTGCGTAGCTGCCCTTATGTTTGCGGCTCCGGGACGCCCGCTAGTTGGTCGGCGGGGAA
  Y  L  Q  Q  R  I  D  G  N  T  N  A  E  A  L  R  A  I  N  Q  P  P  L

GCGGTGACCACCCACCCGTTTGATACTCGCCGCGGTGAATTGGCCCCCGCAGCCCGTAGGTTAGGACGCTC  1120
CGCCACTGGTGGGTGGGCAAACTATGAGCGGCGCCACTTAACCGGGGGCGTCGGGCATCCAATCCTGCGAG
  A  V  T  T  H  P  F  D  T  R  R  G  E  L  A  P  Q  P  P  V  G
```

FIG. 1
(Page 4 of 4)

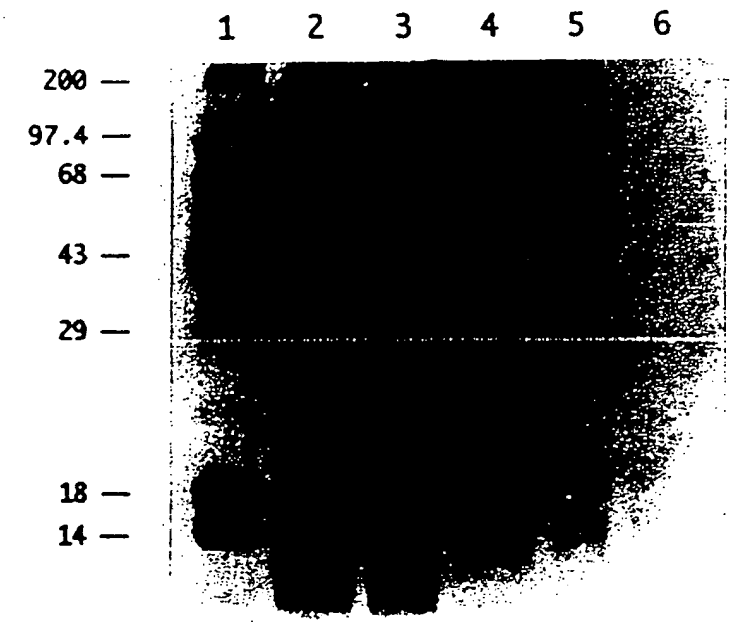

FIG. 2

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/10375

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.
US CL :435/6, 7.1, 240.2, 320.1; 514/2; 530/300, 387.1; 536/23.7, 24.32
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 7.1, 240.2, 320.1; 514/2; 530/300, 387.1; 536/23.7, 24.32

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

STN: Medline, Biosis, CAPlus, WPIDS, JAPIO, PATOSEP, PATOSWO; APS
search terms: mycobacterium tuberculosis, peptide, polypeptide, protein, epitope, antigen, immunostimulat?, membrane, surface

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	TOMMASSEN et al. Use of the enterobacterial outer membrane protein PhoE in the development of new vaccines and DNA probes. Intl. J. Microbiol. Virol. Parasitol. infect. Dis. 1993, VOL. 278, pages 396-406.	1-26
Y	JANSSEN et al. Immunogenicity of a mycobacterial T-cell epitope expressed in outer membrane protein PhoE of Escherichia coli. Vaccine. 1994, Vol.12, pages 406-409.	1-26
Y	Lim et al. Identification of Mycobacterium tuberculosis DNA sequences encoding exported proteins by using phoA gene fusions. J. Bacteriol. January 1995, Vol.177, pages 59-65.	1-26

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
09 OCTOBER 1996	28 OCT 1996
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer ROBERT A. SCHWARTZMAN Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*